Chapter 1

Literature Review: Current Issues and Proposed Directions for Measuring Emotional Intelligence (EI)

New socio-emotional constructs emerging from the fields of emotional, practical and social intelligence bring new challenges for assessment, particularly in test construction and scoring. This is primarily because portrayals of social and emotional situations are much more difficult to clearly interpret, judge, or manipulate than the stimuli commonly used in conventional cognitive testing. For this reason, scoring algorithms are currently more often based on group judgment rather than on systematic theory or bodies of knowledge – i.e., a consensus approach rather than a standards-based approach to scoring. On a construct validation level, it is unclear how distinct socio-emotional abilities are from the more well-known and well-researched intelligence and personality constructs, or whether possible method effects from idiosyncratic scoring mechanisms obscure these relationships. There are thus two broad goals of the current research. Firstly, two new emotional intelligence (EI) tests are created to allow standards-based as well as consensus scoring: (1) the Situational Test of Emotional Understanding (STEU); and (2) the Situational Test of Emotion Management (STEM). Test construction and scoring are based on theory and methodology not often applied in this field: appraisal theory for the STEU, and the Situational Judgment Test (SJT) paradigm for the STEM. Secondly, these research tools are used to examine the construct validity of EI as a new intelligence, including any possible method effects of scoring differences on EI's construct validity. EI's distinction from both crystallized (Gc) and fluid intelligence (Gf), and from broad and narrow aspects of personality is examined, as is its incremental validity in predicting appropriate indices of success. The hierarchical four-branch model of EI (Mayer & Salovey, 1997) was selected as the theoretical basis for this research. Instruments were constructed to assess the two highest branches of the model: *Understanding* (understanding the sequences and combinations of emotions as well as relating

emotions to situations) and *Management* (regulation and management of emotions to moderate negative and enhance positive emotions).

1. Outline of Chapter 1

This chapter briefly outlines the history, theory, major paradigms, and empirical findings in EI research. Assumptions and potential problems with current assessment methodology are presented, followed by an outline of how this thesis will test these assumptions and ameliorate these problems. Essentially, it is suggested that: (1) standards-based scoring (as opposed to the current consensus-based scoring) is important in interpreting the meaning of the EI construct; and (2) the effects of scoring type and response-option type on EI's relationship with criterion variables such as personality and intelligence need to be tested. An outline of how the STEU and STEM will be developed and used to address these concerns is provided. This chapter concludes with a summary of the current validity evidence for the EI construct, as based on existing measures of EI. This validity evidence is provided both as a theoretical background on what EI *should* relate to if it is to be a valid construct (and hence which criteria are used as validity indicators in the following chapters), and to deduce whether current operationalisations of EI *do* delineate a valid construct.

2. Background to EI: History, Models, and Concepts

EI began to be studied scientifically only in the past 15 years. The first psychological publication was in a relatively obscure journal in 1989 (Salovey & Mayer, 1989). In 1995, Daniel Goleman's book *Emotional Intelligence* both echoed and substantially added to a popular interest that fast grew out of proportion to the knowledge accumulated in only a half a decade by a few academic researchers. The immense popularity of EI has in some ways hindered the field, particularly in terms of conceptual clarity – the sheer number of models and definitions of "emotional intelligence" has meant that vastly different constructs exist under the same label, leading to different and sometimes conflicting claims about what EI predicts.

In an effort to bring order to the field Mayer, Salovey, and Caruso (2000a) categorised models of EI into two types: (1) EI as a form of intelligence, involving cognitive processing of emotional information (referred to as "*ability models*" of EI); and (2) EI as partly or wholly a personality-like trait, or behavioural disposition (referred to as "*mixed models*" EI). An alternative and overlapping conceptualisation of this bifurcation was proposed by Petrides and Furnham (2001), who categorised instruments according to their measurement approaches (rather than theoretical bases) into *ability-EI* (also referred to as performance-based or information-processing EI, see Petrides & Furnham, 2000) and *trait-EI*.[1] Generally, instruments following an ability-based theoretical model use an ability-based method of measurement (i.e., involve questions where knowledge or skill are assessed by maximal performance) and instruments following a mixed model conceptualisation use self-report or other-report rating scales. However, this characterisation is somewhat imperfect – some self-report scales are constructed after ability-based models. In this thesis, instruments are classified as trait EI or performance-based EI according to Petrides and Furnham's division by measurement approaches (self-report versus ability scales), as correlational evidence suggests the measurement method defines the construct far more than the underlying theory (e.g., Bastian, Burns, & Nettelbeck, 2005).

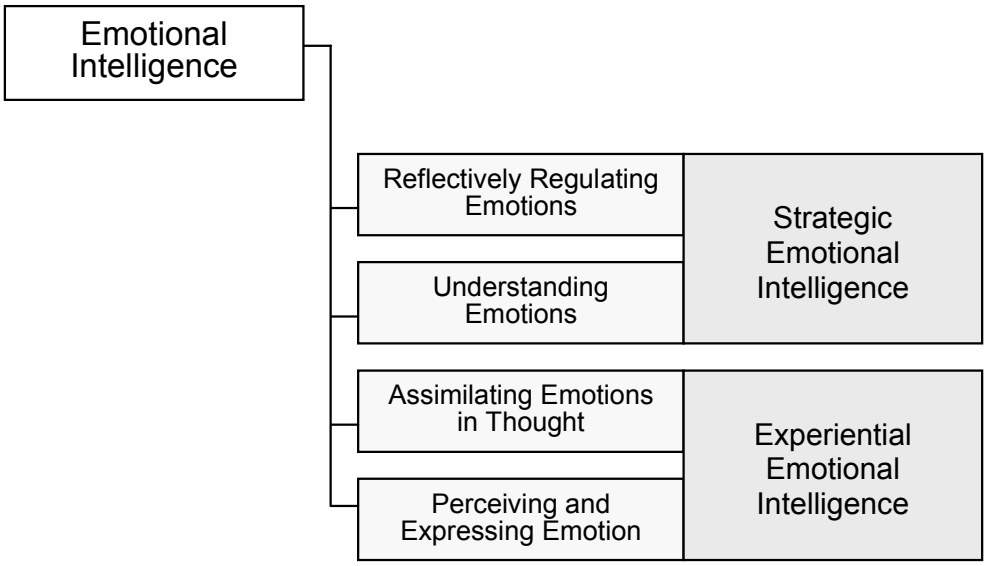
The most commonly agreed-upon definition of emotional intelligence comes from a performance-based model of EI: the four-branch hierarchical model shown in Figure 1.1 (see e.g., Mayer & Salovey, 1997; Mayer, Salovey, Caruso, & Sitarenios, 2001). Under this conceptualisation EI consists of four branches of abilities that increase in complexity from the first to fourth branch. The component abilities in the higher branches depend or build on abilities in the lower branches. At the simplest level (Branch 1), EI is the perception and expression of emotions (*Perception*). Branch 2 consists of the integration of emotions in thought processes (*Facilitation*). Branch 3 includes the

[1] The term "trait EI" can be somewhat awkward for research psychologists given that structural equation modelling uses the term "latent trait" to define hypothetical entities that may be cognitive, dispositional or even representative of other types of constructs, such as biodata. However, in EI research the term "trait EI" is used to define EI as a personality trait as opposed to a cognitive ability, rather than indicating a latent construct.

understanding of emotion labels, relations between emotions, between emotions and circumstances and transitions among emotions (*Understanding*). Finally, the highest branch involves managing emotions in order to moderate negative, and enhance positive, emotions (*Management*). The first two branches are collectively defined as *Experiential EI* (expression, perception, and generation of emotions) and the last two as *Strategic EI* (the understanding and management of emotions; Mayer, Salovey, & Caruso, 2002). This division is supported empirically by factor analyses of instruments based on this model (e.g., Mayer, Salovey, Caruso, & Sitarenios, 2003).

Figure 1.1

The Four-Branch Hierarchical Model of Emotional Intelligence (Adapted from Mayer, Caruso, and Salovey, 2000b, p. 269)



This theoretical model has formed the conceptual basis for some of the self-report (or "trait") measures of EI, either partly (e.g., Petrides & Furnham's [2003] Trait Emotional Intelligence Questionnaire [TEIque]) or wholly (Schutte et al.'s [1998] Self Report Scale [SSRI], which was based on an earlier definition of EI that essentially included only branches 1, 2, and 4). However, the

instruments most strongly modelled on this theory are the performance-based Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al., 2002a) and its non-commercial precursor, the Multi-Factor Emotional Intelligence Scale (MEIS; Mayer, Salovey, & Caruso, 1997).

3. Evaluation of Self-Report EI

As well as the hierarchical four-branch model of EI, there are also a large (and growing) number of conceptual models best classified as mixed models or as trait EI. The primary challenge to validity for trait EI models is their overlap with personality and lack of relation to intelligence. Many researchers argue that EI ought to be part of the intelligence rather than the personality domain to justify the label "emotional *intelligence*" (Matthews, Zeidner, & Roberts, 2002; Mayer et al., 2001; Roberts, Zeidner, & Matthews, 2001). However, the view that trait EI may be a more fruitful direction for EI research has also been expressed (e.g., Perez, Petrides, & Furnham, 2005). In line with the "EI as an intelligence" argument, EI should be conceptually and empirically distinct from personality with relations to personality similar to other cognitive abilities.

One of the most prevalent theories of personality is the five-factor model. This theoretical model is used in the majority of the studies examining the relationship between trait EI and personality. Briefly, the five dimensions of personality are: (1) Neuroticism/Emotional Stability (N); (2) Extraversion/Introversion (E); (3) Openness to Experience (O); (4) Conscientiousness (C); and (5) Agreeableness (A; Costa & McCrae, 1991; Goldberg, 1990). These five broad personality dimensions were later elaborated to include six narrow facets of personality underlying each dimension (Costa & McCrae, 1995). This model is generally known by the name of the primary instrument measuring it: the Neuroticism Extraversion Openness Personality Inventory Revised (NEO-PI-R), after the original version containing only N, E, and O; Costa & McCrae, 1980, 1992). This thesis employs the five-factor model and the NEO-PI-R instantiation of this model as the framework for personality.

Cognitive abilities relate only trivially to personality traits (i.e., less than $r = .10$; Cohen, 1988) for all the five factor dimensions except Openness. Openness relates to crystallized intelligence

(Gc) at about $r = .30$, but only trivially to fluid reasoning ability (Gf; see Ackerman & Heggestad, 1997; also Ashton, Lee, Vernon, & Jang, 2000). Thus EI-personality relations should be in this range if EI is to meet the criteria for an intelligence. Even given that EI might be considered part of the personality domain (Petrides & Furnham, 2001, 2003), the construct must still be distinct from existing personality factors, with correlations to known traits less than $r = .50$ (Cohen's definition of "large").

There are two major mixed-model conceptualisations of EI: (1) Bar-On's (2000) Social and Emotional Intelligence; and (2) Boyatzis, Goleman, and Rhee's (2000) Emotional Competence, which grew out of Daniel Goleman's work on EI (see Goleman, 1995, 1998). Bar-On's theoretical model forms the basis for the Emotional Quotient Inventory instrument (EQ-i), and Boyatzis et al.'s theoretical model forms the basis for the Emotional Competence Inventory instrument (ECI-360). Bar-On's model (and instrument) consists of five broad factors of EI, each composed of several narrower facets: Intrapersonal, Interpersonal, Stress Management, Adaptability, and General Mood. The TEIque (Trait Emotional Intelligence Questionnaire) covers the content domain of this model as well as Mayer and Salovey's (1997) four-branch hierarchical model described previously (Petrides & Furnham, 2003). Emotional Competence was designed to cover five theoretical clusters of more specific abilities; however, empirical analyses later suggested a four-cluster model: Self-Awareness, Self-Management, Social Awareness, and Social Skills (Boyatzis et al., 2000).

There is clear theoretical overlap between these models of EI and personality: The EQ-i includes assertiveness, happiness, stress tolerance, and optimism, while the ECI includes trustworthiness, conscientiousness, and achievement orientation, all of which are either narrow facets or broad dimensions in the NEO-PI-R personality model (Costa & McCrae, 1995). This conceptual correspondence of self-report EI with personality is reflected in empirical findings, even when self-report measures are based on more ability-like frameworks. For example, the SSRI correlates between $r = -.28$ to $r = -.64$ with Neuroticism and greater than $r = .50$ with Openness to Experience and Extraversion (Engelberg & Sjoberg, 2004; Saklofske, Austin, & Minski, 2003; Schutte et al.,

1998). The TEIque correlates r = -.70 with Neuroticism and r = .68 with Extraversion (Petrides & Furnham, 2003). An assessment of the clarity with which one experiences emotions, pays attention to emotions, and tends to repair negative emotions – the Trait-Meta-Mood Scale (TMMS; Salovey, Mayer, Goldman, Turvey, & Palfai, 1995) – is also worth mentioning here. It relates strongly to personality, with repair and clarity subscales correlating around r = .50 with Agreeableness and Neuroticism (Davies, Stankov, & Roberts, 1998).

It is clear from these results that trait EI is part of the personality domain and that some measures may be substantially replicating existing dimensions of the five factor model of personality. Correlations with narrower facets of personality are even higher, indicating substantial overlap if not identity. For example, EQ-i scores correlate greater than r = .75 with the anxiety facet of Neuroticism (Newsome, Day, & Catano, 2000; O'Conner & Little, 2003). However, studies examining the relationship between trait EI and personality have tended to stop at the five broad dimensions rather than the narrower 30 facets (e.g., Austin, 2004; Austin, Saklofske, Huang, & McKenney, 2004; Furnham & Petrides, 2003; Saklofske et al., 2003). Petrides and Furnham (2001) provide an exception to this rule and Petrides and Furnham (2003) measured personality at the facet level but did not report facet-level information. EQ-i relations to anxiety suggest that facet-level as well as broad dimension-level relationships are necessary to confirm the distinctiveness of EI from personality.

Results from the trait-EI field suggest two directions for this thesis. Firstly, to measure EI as a cognitive ability, self-report scales should not be used. For this reason, the STEU and STEM measures are constructed as ability scales rather than self-report inventories. Secondly, when assessing the distinctiveness of trait EI measures from existing personality constructs, it is vital to examine narrow conceptualisations as well as broad domains of personality. This has only rarely been done in the past, and comparing both the newly constructed performance-based tests of EI and an existing trait EI measure (the SSRI) to the 30 narrow facets of the NEO-PI-R model of personality is undertaken in Chapter 5 of this thesis.

4. Performance-Based EI: Methodological Issues

The MSCEIT and its earlier manifestation (the MEIS) are virtually the only performance-based tests of EI and thus almost all that is known empirically about EI is based on studies using these instruments.[2] Debate regarding the validity of performance-based EI as a new intelligence has centred on criteria of common intelligence-related and unique EI-related variance. Mayer et al. (2001) call this the correlational criteria – that test scores should be "similar to, but distinct from, mental abilities described by already established intelligences" (p. 270; see e.g., Kaufman & Kaufman, 2001; Mayer, Caruso, & Salovey, 2000b; Mayer, et al., 2001; Roberts et al., 2001 for some examples of this debate). Given that the MSCEIT uses both idiosyncratic scoring rubrics *and* an unusual response-format for a cognitive ability test, it may be that the unique variance identified in studies to date relates to method-of-measurement variance as much as construct-related variance (i.e., the distinctiveness of scoring and response format, rather than the distinctiveness of EI may distinguish the MSCEIT tests from other measures of intelligence). A conceptual examination of the MSCEIT's method of measurement is thus needed and is described below.

4.1 Response Format

The MSCEIT's eight tests are described in Table 1.1, along with a detailed description of the component abilities of the four branch model (taken from Mayer & Salovey, 1997; and Mayer et al., 2002a, 2002b). All the tests from branches 1, 2, and 4 have a response format of ratings-based judgments: The presence of emotion, helpfulness of a mood, similarity of emotion to sensation, or effectiveness of a response to a situation are each rated on five-point scales.[3] Tests from the third

---

[2] There are at least two other performance-based tests of EI, but both of these assess a single construct, rather than the entire content domain of the four-branch hierarchical model of EI: the LEAS (Levels of Emotional Awareness Scale; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990) and the EARS (Emotional Accuracy Research Scale, Mayer & Geher, 1996; Geher, et al., 2001). In addition, there are several measures of emotion recognition in faces or other stimuli, but these are not generally considered part of the EI field, and therefore research in these has not informed knowledge about EI until very recently.

[3] Ratings-based responses are also used in practical and social intelligence tests (e.g., Hedlund et al., 2003; Legree, 1995; Sternberg et al., 2000, Weis, Seidel, & Suess, 2005).

branch of the MSCEIT (Emotional Understanding) are multiple-choice rather than ratings-based, as tests of cognitive ability more traditionally are. Two of the MEIS' four Understanding tests are also multiple-choice, whereas the rest of the battery asks for ratings-based judgments. Ratings-based judgment formats may be problematic for two reasons: (1) there may be individual differences in the use of the rating-scale, and (2) the unusual response format may act as method-related variance in relationships between EI branches, and between EI and intelligence. Both these issues are examined empirically in the current thesis, with Chapter 3 examining both issues directly and a comparison of subsequent chapters providing more indirect tests of these possible method effects.

Ratings-based formats may be problematic if response bias toward one end of the scale affects responses (Legree, 1995). For example, particularly cautious or defensive people might use only the lower end of the scale whereas optimistic or suggestible people might rate all options quite highly: Meaningful individual differences in the use of the ratings-scale might thus affect scores on the test. Legree suggests profile scoring be used to control for individual differences in rating-scale use. His version of profile scoring standardised all ratings over each participant prior to scoring items, but Pearson or Spearman correlational scoring are other alternatives (Legree, Psotka, Tremble, & Bourne, 2004). Mayer et al. (2000b) found no gender differences in rating-scale use for the MEIS' Stories test, but did not examine other tests or other sources of individual differences. Chapter 3 of this thesis examines individual differences in rating-scale use in more detail, looking at gender, cultural, and personality-related differences as sources of variation in rating-scale use. Profile scoring of ratings-based tests will be undertaken if there is evidence of individual differences in rating-scale use.

Table 1.1

Structure of the Four-branch Hierarchical Model of EI, and Descriptions of Tests from the Mayer-Salovey-Caruso Emotional Intelligence Test Battery (MSCEIT), at Each Branch

Branch and Definition	*MSCEIT Test and Description*
AREA 1: EXPERIENTIAL EMOTIONAL INTELLIGENCE	
Branch 1: Emotion Perception	
• Identify emotion in physical states, feelings, and thoughts • Identify emotions in other people, designs, and artwork, through language, sound appearance or behaviour • Accurately express emotions and needs relating to emotions • Discriminate between accurate or inaccurate, and honest or dishonest emotional expression	*Faces*: For each face, rate the presence of five emotions of from 1 (no emotion) to 5 (extreme emotion). *Pictures*: For each picture, rate the presence of five emotions of from 1 (no emotion) to 5 (extreme emotion).
Branch 2: Emotions Facilitate Thought	
• Emotions prioritise thinking by directing attention • Emotions generated to aid judgment and memory • Emotions change the individual's perspective, allowing multiple points of view • Emotional states differentially encourage different problem solving approaches (e.g., happiness facilitates inductive reasoning)	*Sensations*: For each scenario, rate the helpfulness of three different moods from 1 (not useful) to 5 (useful). *Facilitation*: For each statement, make three judgments on the similarity of an emotion to a physical sensation from 1 (not alike) to 5 (alike).
AREA 2: STRATEGIC EMOTIONAL INTELLIGENCE	
Branch 3: Emotional Understanding	
• Label emotion and recognise relations among these labelled emotions. • Interpret meanings of emotions and circumstances • Understand complex feelings • Understand transitions among emotions	*Blends*: Multiple-choice questions ask which emotions are related to particular situations. *Changes*: Multiple-choice questions ask which combinations of emotions form which complex emotions
Branch 4: Emotion Management	
• Stay open to feelings, both pleasant and unpleasant • Reflectively engage or attach to emotion depending on its utility or informativeness • Reflectively monitor and judge emotions in self and others • Manage emotion in oneself and others: moderating negative and enhancing positive emotions	*Management*: For each scenario, rate the effectiveness for mood-management of four actions from 1 "very ineffective" to 5 "very effective" *Relations*: For each scenario, rate the effectiveness of three responses from 1 "very ineffective" to 5 "very effective".

This ratings-based format may be responsible for EI's distinctiveness from other types of intelligence (see, e.g., Ciarrochi, Chan, & Caputi, 2000; Davies et al., 1998 for evidence that EI is

distinct from existing measures of intelligence). The stronger relationship between intelligence and EI found for Understanding than for other branches may very well be due to the response format (multiple-choice) which is different to that of the rest of the MSCEIT (where tests are ratings-based). Research thus far has found that Understanding is the most "cognitively saturated" of the EI branches, with Understanding tests differentiated from other EI tests by their higher correlations with intelligence (Mayer et al., 2001). When the link between EI and intelligence is examined at different branches, Understanding tests are either the only branch of EI that correlates significantly with intelligence, or correlate much more strongly than tests from the other branches (Barchard, 2003; Bastian et al., 2005; Lam & Kirby, 2002; Lopes, Salovey, & Straus, 2003; Mayer et al., 2000b; O'Conner & Little, 2003; Roberts et al., 2001; Warwick & Nettelbeck, 2004).

This thesis examines whether response-format affects relationships between different branches of EI and between EI and intelligence or personality. The Situational Test of Emotion Management (STEM) was developed in both ratings-based and multiple-choice forms. By way of illustration, a hypothetical example of the same item in these two different formats is given in Table 1.2. Chapter 3 compares the relationship of each form to other measures of EI (both ratings-based and multiple-choice), personality, and a measure of intelligence. Particular attention is paid to whether Understanding and Management actually form two separate constructs when both are measured with items of the same response format.

Table 1.2

Example of the Same Item in Multiple-Choice and Ratings-Based Format

Multiple-choice	*Rating scale for each option*					
It is raining and an office worker without an umbrella must get to her car. *The most effective action for her is to?*	It is raining and an office worker without an umbrella must get to her car. *How effective are each of the following actions from 1 (not at all effective) to 5 (extremely effective)?*					
		Not at all effective				Extremely effective
(a) Steal someone else's umbrella.	(a) Steal someone else's umbrella.	1	2	3	4	5
(b) Run to the car in the rain.	(b) Run to the car in the rain.	1	2	3	4	5
(c) Wait for the rain to stop.	(c) Wait for the rain to stop.	1	2	3	4	5

4.2 Scoring

In contrast to traditional aptitude or cognitive tests such as arithmetic, reasoning, spatial skills or vocabulary, where organised systems such as mathematics, logic, geometry or semantics determine the correct answer veridically, there are no equivalent systems for determining the correct answer to tests assessing EI, particularly in the strategic area (Roberts et al., 2001). This is because items from the strategic area (especially emotion management) present realistic situations that can mirror the complexity of real-life situations. With such complexity, unambiguously applying a set of specifications or rules about emotion management would be difficult, even if such a set of rules were contained within psychological theories. The MEIS and MSCEIT do *not* use theoretical systems as the basis for scoring, even for emotion perception, where this is demonstrably possible (e.g., Ekman, 2004; Matsumoto et al., 2000; Scherer, Banse, & Wallbott, 2001). Instead, solutions to the problem of the correct answer are undertaken in three different ways: (1) target-scoring, (2) expert-scoring, and (3) consensus-scoring. A discussion of these scoring rubrics follows.

4.2.1 Target-scoring

Target-scoring occurs when the creator or target of the item stimuli determines the correct answer: e.g., a person photographed expressing a particular emotion will decide on the correct

interpretation of their facial expression. This is obviously only possible for items where there is a clear target and is only used for some tests of the MEIS (the earlier version of the MSCEIT). The EARS (Emotional Accuracy Research Scale; Mayer & Geher, 1996; Geher, Warner, & Brown, 2001) and the Manheim University Social Intelligence Test (Weis et al., 2005) also use target-scoring.

4.2.2 Expert-scoring

Expert-scoring occurs when an expert (or panel of experts) determines the correct answer. Both the MEIS and MSCEIT use this as one possible form of scoring. The MEIS uses two of the three test authors as experts and the MSCEIT (Version 2) uses 21 members of the International Society for the Research of Emotions as experts (Mayer, et al., 2003). In some ways this simply pushes the specification of the correct answer one step back, into the specification of criteria for expertise. Certainly academic study of the emotions would lead to greatly enhanced emotion-related knowledge or any declarative aspects of EI and thus would seem a reasonable definition of expertise. However, Roberts et al. (2001; see also Zeidner, Matthews, & Roberts, 2001) propose that there might be multiple domains of expertise, of which academic knowledge is only one. Other domains might include experience in a profession geared towards emotional healing (e.g., counselling, psychiatry, psychotherapy, some forms of religious leadership), or towards understanding and managing people's relationships and goals (e.g., human-resource-related careers, coaching). Adding to the importance of expert selection procedures in EI is the evidence that scores are higher for test-takers similar to the experts – white males scored more highly under expert scoring for the MEIS when the experts were white males (Roberts et al., 2001).

Given that the STEM test developed in this thesis uses expert scoring as one possible scoring mechanism, such concerns are highly relevant. Care was taken to select experts from multiple domains (rather than academic research only). In addition, Chapters 2 and 5 examine whether the gender composition of the expert samples affects group differences on tests scored by these experts' judgements (i.e., whether males do better when the expert screening sample is all male and females do better when all the experts are female).

4.2.3 Consensus scoring

The MEIS and MSCEIT are most commonly scored by consensus. Legree et al. (2004) outline the logic of consensus scoring, presenting it as an extension of expert scoring. Legree et al. reason that the judgment of experts is equivalent to non-experts (on tasks of tacit or social knowledge) except that they are less consistent and therefore less reliable. Non-expert judgments can be considered as composed of two parts: the common variance (expert opinion) and unique variance, which ought to be random. If this is so, an aggregate of non-expert judgment is equivalent to expert judgment.

For the MSCEIT, both expert and consensus scores are awarded according to a proportional algorithm. Normative or expert samples first complete the test and the proportion of these samples choosing each option becomes the score that is awarded for that option (e.g., if 61% of the screening sample selected a rating-scale point of "1" for happiness in a face, a response of "1" would get a score of 0.61). However, this proportion-based algorithm is not the only way that expert or consensus judgment might be operationalised. Indeed, in ratings-based items where several ratings of one stimulus are elicited (e.g., ratings of happiness and sadness in the same face) considering these as one item introduces experimental dependence, thereby artificially inflating inter-item correlations and internal consistency estimates (MacCann, Matthews, Zeidner, & Roberts, 2003). Using Euclidean distance scoring (or its more complex derivative, Mahalanobis scoring, which adjusts for the co-variation in each testlet) gets around this problem by treating all responses based on the one stem as one multi-part item. This technique is used currently in tacit knowledge research, resulting in reasonable reliability estimates (Cronbach's alpha = .66 to .76; Hedlund et al., 2003). As discussed earlier, profiles of the multiple ratings of each stimulus would be even more appropriate if there are meaningful individual differences in the use of the ratings scale. For both reasons of individual differences in rating scale use, and experimental dependence, ratings-based tests constructed in this thesis will be scored according to distance and profile scores only (and not proportionally).

A further weakness with the MSCEIT's proportional consensus scoring is that distributions of total test scores cannot logically be both normally distributed and internally consistent. In any given question, the majority will score the highest mark. If the tests are to be internally consistent, then subsets from the same group of "high ability" people (who form a majority) will answer the "best" option on most items and skew at the item level will be accentuated at the total score level. The distribution of test scores will be highly negatively skewed and leptokurtic (i.e., most scores will form a highly peaked cluster at the top end of the distribution). Proportion-based consensus scoring thus inevitably results in a trade-off, with a reliable test with a high level of skew and kurtosis on one hand, or an unreliable test with a reasonable distribution shape on the other (MacCann, Roberts, Matthews, & Zeidner, 2004). It is uncertain whether this effect would also be apparent with expert proportional scoring, but if the expert option proportions are similar to the group consensus option proportions then this effect would also be apparent for expert scores. It is also uncertain whether this effect would be apparent for multi-dimensional (profile or non-profile) scoring of ratings-based tests.

One obvious concern is that scores created by different methods should not be too dissimilar – tests scored in one way should clearly be measuring the same construct as tests scored under a different rubric. MSCEIT consensus scores are strongly related to MSCEIT expert scores ($r = 0.93$ to $r = 0.99$; Mayer et al., 2001), indicating that both scoring methods index the same construct. However, this convergence of scoring was not true of the earlier MEIS, particularly for Emotion Perception branch scores; where some correlations were in fact significantly negative (Roberts et al., 2001). One of the reasons for this discrepancy is probably that expert scores for the MSCEIT were proportion scored (i.e., the proportion of experts selecting each response was the score awarded) whereas MEIS expert scores were dichotomously scored (with 1 point awarded to the rating-point selected by the 2 experts, as well as the rating-points on either side).

Consensus scoring may also be conceptually problematic, particularly for ratings-based questions. If correctness is equivalent to agreement with the majority, those with exceptionally high skill at (for example) facial expression recognition may be able to perceive nuances of expression the

ordinary person would not. Such people would disagree with the majority and score very poorly on a consensus-scored test. Given the conceptual and empirical difficulties with target, expert, and consensus scoring, a primary goal of this research program was to examine the possibility of *standards-based scoring* (i.e., determining the correctness of a response based on theory, knowledge, or empirical evidence). Standards-based scoring might be expected to be similar to expert-based scoring (as, by definition, experts are expected to be correct).

One last potential issue with consensus scoring is that it is not used in tests of intelligence. As with the ratings-based format of EI tests, this difference in methodology between tests of EI and other intelligence tests may at least partly be responsible for the empirical distinctiveness of EI. Chapter 5 of this thesis examines this issue in some detail. The distinction of consensus-scored EI tests from both conventionally-scored and consensus-scored intelligence tests is determined. If consensus scoring is acting as a method effect that distinguishes EI from intelligence, then correlations between EI and intelligence should be higher when intelligence tests are scored by consensus.

5. Dealing with Methodological Issues: Standards-Based Scoring and Test Construction Approaches

Scoring items according to systematic rules or formal bodies of knowledge is referred to as standards-based scoring (i.e., scores are based on some clear verbalisable standard rather than human judgment). In contrast to the strategic area of EI, items from the experiential area (particularly emotion perception) could be argued to have systematic rules for correctness that might allow standards-based scoring of these tests. These rules would be analogous to systems such as logic or geometry used to score reasoning or spatial skills tests, or bodies of knowledge such as a dictionary used to score vocabulary tests. For example, Paul Ekman has led an enormous amount of work on facial expressions of "basic emotions", finding cultural invariance in the expression and recognition of these emotions (e.g., Ekman, 1992; Ekman et al., 1987; Ekman & Keltner, 1997). Indeed, specific rules relating muscular changes to facial expressions were codified decades ago in the form of the Facial Action Coding System (FACS; Ekman & Friesen, 1978; Ekman & Rosenberg, 1997). More

recently, individual differences measures and associated training tools have been created out of this research (e.g., Matsumoto et al.'s JACBART [Japanese and Caucasian Brief Affect Recognition Test], 2000; and Ekman's Micro-Expression Training Tool [METT] and Subtle Expression Training Tool [SETT], 2004). Similarly, a measure of emotion recognition in vocal expression with definitive right and wrong answers was recently created (Banse & Scherer, 1996; Scherer et al., 2001). However, such standards-based measures have only recently begun to be used in EI research (Austin, 2004; Roberts, Schulze, O'Brien, Reid, MacCann, & Maul, submitted).

For the sake of construct validity, a test's answers should depend on known facts about the universe and not solely on group opinion. To the author's knowledge, there are currently no tests of Strategic EI with standards-based scoring systems. However, there is empirical evidence on both the antecedents of emotion generation and the differential effectiveness of different coping strategies for different types of situations that might be used as the basis for such scoring. Appraisal theories relate specific combinations of appraisals to the generation of specific emotions. These theories could be used to create item stimuli where particular emotions are present or absent according to these rules (i.e., to create and score items assessing knowledge of situational antecedents of emotion). Similarly, a scoring key outlining which strategies are most effective for different types of situations might be derived from some of the coping literature (i.e., a scoring key for emotion management tests). This is precisely the plan in the current research: to construct standards-scored tests of Emotional Understanding and Emotion Management (the two aspects of Strategic EI) based on appraisal and coping research, respectively.

5.1 Development of the Situational Test of Emotional Understanding (STEU)

Understanding emotion is the "ability to comprehend emotional information about relationships, transitions from one emotion to another, [and convey] linguistic information about emotions" (Mayer et al., 2003, p. 235). Mayer and Salovey (1997) describe developing Emotional Understanding as learning the link between classes or categories of situations and the classes or categories of emotions that accompany these (e.g., loss produces sadness; threat produces fear).

Mayer et al. (2003) describe Emotional Understanding as conveying information about relationships between people and other people or objects. The core essence of Emotional Understanding appears to be *knowing which types of situations represent which emotions*, along with the ability to verbally convey this knowledge.

Appraisal theories of emotion provide a set of rules for which types of circumstances act as causes or precursors to which types of emotions and thus can act as scoring criteria in the construction of Understanding tests. Part I of Chapter 2 describes how Roseman's theory of the structure of emotions was used to create and score test items on an Understanding test (see Roseman, 1984; 2001; Roseman, Antoniou, & Jose, 1996; Roseman, Wiest, & Swartz, 1994, for the structure of emotions). This structural theory of emotions describes the combination of appraisal dimensions (e.g., low control potential, self-caused) that cause each of 17 emotions. In the current research, situations were developed for each emotion to represent the combination of appraisal dimensions outlined in the structural theory. Test-takers needed to recognise which emotion would develop from each of these situations. Since there was an a priori correct answer according to an empirically supported theory, this constituted standards-based scoring. This test development process is described in more detail in Part I of Chapter 2.

5.2 Development of the Situational Test of Emotion Management (STEM)

Mayer and Salovey (1997) exemplify emotion management as knowledge of strategies for dealing with emotional situations. Such strategies range from rudimentary control strategies (e.g., counting to 10 when angry) through to sophisticated reasoning involved in manipulating others' emotions (e.g., motivating others to oppose injustice). In test items from the MSCEIT and MEIS, participants rate the effectiveness of strategies for different situations, implying that emotion management is not simply the number of strategies known, but the knowledge of how effective different types of strategies are in different situations. Emotion management is thus defined here as knowledge of which strategies are most effective in managing specific emotional situations.

In the current research, the Situational Judgment Test (SJT) paradigm is used to construct item stems and response options for a test of emotion management. The first step used in developing an SJT is sampling the situational domain via incident generation from a relevant population (Motowidlo, Dunnette, & Carter, 1990; Ployhart & Ehrhart, 2003). Further steps in SJT construction outlined by Ployhart and Ehrhart are: (2) the generation of responses to the situation by a relevant sample; (3) the summarising of these responses into short response options; and (4) the review of items by subject matter experts to determine the correct response. Part 2 of Chapter 2 outlines how the STEM (Situational Test of Emotion Management) was constructed according to this method. In Study 1, emotional situations were collected from a variety of participants and content analysis used to create amalgamated hypothetical situations to use as items. In Study 2, a second sample generated ideal and actual responses to these situations, which were condensed to form response options. The score awarded to each response option was then generated in one of two possible ways: (1) *by experts*, where expertise was taken from a range of possible domains, (2) *by standards*, according to a scoring key developed from a literature review on the differential effectiveness of different coping strategies in different types of situations. The validity and methodology for expert scoring is described in Study 3 and the coping-derived scoring key in Study 4.

6. Validity Issues in Performance-Based EI

6.1 Structural Issues: Four Branches and Two Areas?

In general, neither the MEIS nor MSCEIT have provided support for the four-branch model of EI. Ciarrochi et al. (2000) found a two-factor solution to the MEIS that distinguished between Perception and the other three branches. Roberts et al. (2001) found a three-factor solution to the MEIS where Perception, Management, and Understanding were distinguishable, but no Facilitation factor emerged. Roberts et al. (submitted) found a similar solution to the MSCEIT (with factors for Experiential EI, Understanding, and Management). Palmer, Gignac, Manocha and Stough (2005) found a similar three-factor solution to the MSCEIT for both their own data, as well as data reported in Mayer et al. (2003). Day and Carroll (2004) similarly reported that although they obtained a four-

factor model, Facilitation and Perception branches were almost perfectly correlated at $r = .96$, and Management and Understanding factors were very strongly correlated at $r = .84$. In Mayer et al.'s four-factor solution, Management and Understanding factors strongly correlated strongly at $r = .76$. However, this four-factor solution could only be obtained by forcing the correlation between the first two branches (Perception and Facilitation) to be equal to that between the last two branches (Understanding and Management).

In addition, Roberts et al.'s (2001) Understanding factor was defined only by the two tests with multiple-choice response format. The two ratings-based tests of Understanding loaded more highly on the *Management* factor, supporting the possibility that a response-format method factor may be responsible for reported factor structures. Mayer et al. (2000b) found the same three factors, although all four of the Understanding tasks and the two Facilitation tasks loaded on the Understanding factor (with the two highest loadings for the multiple-choice tasks).

In sum, there is relatively clear support for the delineation of EI into an experienced-based component (the experiential area) and a component involving reasoning and use of strategies (the strategic area) but there is *not* particularly strong support for the further division of these areas into the four hypothesised branches. This thesis concentrates primarily on the strategic area, and the differentiation of Understanding and Management.

There are several reasons why this thesis focused on Strategic EI in particular. Firstly, several standards-scored instruments of Perception already exist (often also known as *Emotion Recognition Ability* measures, see Matsumoto et al., 2000; Roberts et al., submitted). Emotion perception and particularly facial expression recognition are well-researched and relatively well understood, allowing the development of such standards-scored measures (O'Sullivan & Ekman, 2004). Given that Facilitation does not emerge as a factor, and behaves differently in different analyses, it seems that Experiential EI is defined primarily by Perception, and as such is well-represented by existing measures. Secondly, major models of intelligence, while including perceptual abilities, focus on knowledge and reasoning as key components of intelligence (e.g., Carroll, 1993; Horn & Noll, 1997).

The structure of the four-branch hierarchical model of EI suggests that EI also follows this trend: Although lower-level information processing of emotional stimuli via perception is necessary, higher-level reasoning and knowledge form the apex of the hierarchical model, and are the most cognitive of the components (Mayer et al., 2001; Roberts et al., 2001). It is at this level that proof of the distinction between EI and existing intelligences is most needed (discussed in further detail in Section 5.2.1).

The structural question examined in Chapters 3 and 4 is whether Understanding and Management branches are distinct, particularly when tests from both branches are given in the same response format. Given very high correlations between these abilities, it is possible that the distinction between Understanding and Management is dependent on their different response formats in the MSCEIT. Chapter 5 examines whether Strategic EI is in fact distinct from existing measures of intelligence.

6.2 Construct Validity of Performance-Based EI

As previously discussed, EI measures must relate to but be distinct from existing measures of intelligence if EI is to qualify as a new type of intelligence. They should also be independent of personality dimensions at both the broad and narrow level. As Landy (in press) points out, discriminant validity evidence is useless in the absence of any criterion prediction or evidence of convergent validity. One of the possible advantages that EI has over traditional measures of intelligence is the prediction of achievement outside formal academic and workplace environments, in areas of life where emotion understanding and management may be more important than reasoning or formal knowledge (e.g., Bastian et al., 2005; Roberts, Schulze, Zeidner, & Matthews, 2004; Saarni, 1999). Examples of such areas examined are many and varied, including (lack of) deviant or aggressive behaviour such as fighting, excessive use of alcohol and other drugs, marital satisfaction, and use of positive coping strategies. Existing evidence for these relationships between the MEIS/MSCEIT and these constructs (intelligence, personality, and emotion-related criterion variables) is described below.

6.2.1 Validity of Performance-Based EI: Relationships with Intelligence

Existing conceptions of intelligence may be either single-faceted (general intelligence or psychometric *g*; see e.g., Jensen, 1998) or consist of numerous group factors. When using relationships to intelligence as a criterion for EI's validity, group factor theories may prove more useful for several reasons (Matthews et al., 2002). EI's distinction from *each* narrow component of intelligence must be assured for EI to constitute a new ability. Distinction from intelligence at the broadest level only does not exclude the possibility that EI is similar or identical to a narrower component within the structure of mental abilities. Just as the EQ-i needed to be differentiated from the narrow components of Neuroticism, so performance-based measures must be differentiated from narrow components of cognitive ability in order to achieve the status as a new construct. Secondly, using a multiple-ability framework is useful in that existing knowledge about age progressions, neural pathways, group differences, and real-world predictions of different group factors might inform research on EI.

Two commonly accepted group factor models for the structure of intelligence are the theory of fluid and crystallized intelligence (Gf/Gc theory, see e g., Horn & Cattell, 1966; Stankov, 2000) and Carroll's three-stratum model (Carroll, 1993), which are largely equivalent. The major difference is that Gf/Gc theory does not propose an overarching level of general intelligence whereas the three-stratum model does. Briefly, the major agreed-upon dimensions of the nine or more abilities involved in both Gf/Gc theory and three-stratum theory are: (1) fluid intelligence (Gf, or innate reasoning ability using culture reduced material, often hypothesised to be equivalent to *g*); (2) crystallized intelligence (Gc, knowledge due to formal education and acculturation); (3) visualization (Gv; the visual processing and manipulation of information); and (4) auditory ability (Ga, the auditory processing and manipulation of information). Speed and memory factors are slightly different for the two models but can be divided into: (5) speed factors (both cognitive speediness [Gs] and decision speed [Gp] in the three stratum model, but only broad speediness [Gs] in Gf/Gc theory); (6) short-term memory factors (retrieval ability [Gr] in three-stratum theory and short term acquisition and

retrieval [SAR] in Gf/Gc theory); and (7) long-term memory factors (memory and learning [Gy] in three-stratum theory, and tertiary storage and retrieval [TSR] in Gf/Gc theory). Newer research has concentrated on information processing and manipulation in other sensory modalities as possibilities for inclusion at this second stratum level (e.g., Danthiir, Roberts, Pallier, & Stankov, 2001; Roberts, Stankov, Pallier, & Dolph, 1997).

Attempts have been made to unify the two theories and such unification is the basis for the newest incarnation of the Woodcock-Johnson test battery (see McGrew, 1997; McGrew & Woodcock, 2001; the unified theory is referred to as *Cattell-Horn-Carroll or CHC Theory*, after its chief proponents). CHC theory works towards consensus definitions of the broad (stratum II) and narrow (stratum I) abilities empirically explored in Gf/Gc theory and three-stratum theory. The core of CHC abilities are: (1) Gf; (2) Gc; (3) Gv; (4) Ga; (5) Gsm (i.e., short term memory); (6) Glr (long-term retrieval ability); (7) Gs; (8) Gq (quantitative ability), and (9) Grw (reading and writing ability; see McGrew, 2006). In this thesis, intelligence tests are used for two purposes: (1) to determine whether EI meets Mayer et al.'s (2001) correlational criteria for intelligence, and (2) to determine which types of intelligence EI relates to most strongly. For these purposes, it is less important to determine which intelligence theory is definitively correct than to use the common components as the basis for placing EI within such structures.

Examining current findings of the MEIS and MSCEIT through the lens of Gf/Gc theory (with particular reference to Gf and Gc), the broad conclusions are that EI tends to relate to crystallized or verbal measures of ability rather than to fluid reasoning. These relationships are strongest for Understanding tests (Barchard, 2003; Lam & Kirby, 2002; Lopes et al., 2003; Mayer et al., 2000b; Roberts et al., 2001, submitted). The majority of studies examine verbal or crystallized aspects of intelligence, such as vocabulary tests, the ASVAB (Armed Services Vocational Aptitude Battery, see Roberts et al., 2000 for evidence of its correspondence to Gc) or proxies such as verbal SATs or college entrance scores. However, there are a few studies that relate the MEIS and MSCEIT to matrices tests, which arguably measure purely Gf (Bastian et al., 2005; Ciarrochi et al., 2000; Roberts

et al., submitted). Although both Ciarrochi et al. and Roberts et al. found no relationship between EI and matrices test scores, Bastian et al. did find a small relationship between Perception and matrices and a moderate relationship between Management and matrices. In summary, the MSCEIT appears to relate to Gc far more strongly than Gf, with Understanding tests relating particularly strongly.

Given that the strongest relationships with intelligence have been found for Understanding tests (which are multiple-choice) with Gc, it seems that the distinctiveness of EI from intelligence should be examined with multiple-choice EI measures and Gc. Chapter 5 addresses this question, examining whether a Strategic EI factor composed of multiple-choice tests can be distinguished from a Gc factor.

In both MEIS and MSCEIT, almost all tests rely on the correct comprehension and interpretation of written text (with two exceptions where visual interpretation is more important: Faces and Designs). An immediately obvious concern is that correlations with Gc and not Gf area based on variation due to the *ability to read the question* rather than EI-related variation (i.e., EI scores are confounded with the narrow reading comprehension component of Gc). Relations between EI tests, as well as between EI and Gc, might also represent reading comprehension rather than a construct-related shared variation. Both Chapter 3 and Chapter 6 examine this issue. Chapter 3 controls for vocabulary test scores in the relationships between EI tests. Chapter 6 examines the extent that the ability to read the questions accounts for correlations between tests (through reading comprehension and the overlap between verbal and non-verbal measures of EI).

6.2.2 Validity of Performance-Based EI: Relationships with Personality

There have been no strong links with personality found for the MSCEIT, although a small degree of relationship with Agreeableness is consistently reported and is generally stronger for Management than Understanding (Brackett & Mayer, 2003; Brackett, Mayer, & Warner, 2004; Lopes et al., 2004; Lopes et al., 2003; Roberts et al., 2001; Schulte, Ree, & Caretta. 2004). Small but significant correlations have also been found for Openness to Experience, although Lopes et al.'s (2003) findings of a negative correlation contradict this general trend. However, given that

intelligence tasks (particularly Gc) *do* tend to correlate with Openness (in the range of $r = .30$; Ackerman & Heggestad, 1997) and performance-based EI is a form of intelligence, a relation with Openness would make conceptual sense. It is thus expected that the STEU and STEM should show small to moderate correlations in the range of $r = .30$ with Openness to experience and with Agreeableness. In addition, it might be expected that EI would differ from Gc in its relation to Openness at the facet level. Specifically, while Gc relates only to the intellectual/cognitive facets of Openness (Ideas, Values, and Actions; Gignac, Stough, & Loukomitis, 2004)[4], EI might also be expected to relate to the emotional/artistic facets of Openness (i.e., Feelings, Aesthetics, and Fantasy). This relationship is examined in Chapter 5 of this thesis.

6.2.3 Validity of Performance-Based EI: Further Criteria

Workplace performance.

Van Rooy and Viswesvaran's (2004) meta-analysis estimated a correlation of .17 between MEIS scores and workplace performance of ($N = 1368$). By branch, correlations with workplace performance were .13 (for Perception, $N = 3484$); .24 (for Facilitation, $N = 770$); .23 (for Understanding; $N = 1525$) and .19 (for Management; $N = 2961$). Validity coefficients were higher for expert scores than consensus scores (.25 compared to .08 respectively). However, these results were obtained for the MEIS and not the newer MSCEIT, which might be expected to show higher validity coefficients after revisions to content and scoring were implemented.

Substance use and abuse. MEIS scores relate to lower self-reported use and intention to use tobacco and alcohol (Trinidad & Johnson, 2002; Trinidad, Unger, Chou, & Johnson, 2004). The MSCEIT (and particularly the strategic area) predicted self-reported illegal drug and alcohol use (Brackett et al., 2004). Use of better coping strategies may be a possible reason for relationships with less substance use and abuse (e.g., facing or re-framing the problem rather than avoiding or

---

[4] Although Gignac et al. (2004) argued that Openness related to *g* rather than Gc, their *g*-factor had far higher loadings on Gc than Gf tasks and thus was more akin to a Gc than a *g* factor.

ruminating over the problem, and particularly not using destructive methods of avoidance such as substance abuse). Lyons and Schneider (2005) found high Strategic EI related to positive re-appraisal (i.e., challenge rather than threat appraisals), which provides some support for this idea. Chapter 4 of this thesis examines whether Strategic EI is related to tobacco use and frequent alcohol use in adults.

Life satisfaction. MEIS and MSCEIT scores have a small degree of relation to self-reported life satisfaction (Bastian et al., 2005; Ciarrochi et al., 2000; Mayer et al., 2000b). However, Brackett and Mayer (2003) found no relationship between MSCEIT scores and life satisfaction (although MSCEIT scores were related to psychological well-being) and Bastian et al. found this relationship only for Management. Salovey, Stroud, Woolery, and Epel (2002) found a relationship between the MSCEIT and interpersonal satisfaction. The current research examines the relationship of the STEU and STEM to life satisfaction in a different way, separating life satisfaction into current and retrospective components (after factor-analytic evidence from McDonald, 1999). Although relationships to both current and retrospective life satisfaction are expected, relationships to retrospective life satisfaction may be stronger, as this involves an element of positive re-appraisal (a strategy for minimising negative emotions that may be used more often or more effectively by those with high emotional management abilities).

Emotional distress. Ciarrochi, Deane, and Anderson (2002) found that the LEAS (Levels of Emotional Awareness Scale, a performance-based test of emotional awareness possibly most closely corresponding to the MSCEIT's Understanding branch) related to stress. Salovey et al. (2002) and Woitaszewski and Aalsma, (2004) also found relationships between performance-based EI and stress. Bastian et al. (2005) found small relationships between anxiety and MSCEIT Perception and Understanding branches and MSCEIT total scores. After intelligence and personality were controlled, MSCEIT scores predicted an additional 6% of the variance in anxiety. Emotional intelligence may relate to less distress through the better use of coping strategies (e.g., positive re-appraisal of negative events, lack of rumination; Salovey, Bedell, Detweiller, & Mayer, 2000). The current thesis uses

mental health, anxiety, depression, and stress as indicators of state emotional distress acting as criteria by which to assess the validity of the STEU and STEM and the EI construct in general.

Alexithymia. Alexithymia (*a* = lack, *lexis* = word, *thymos* = emotions; Taylor, 2001) was originally conceptualised as a set of symptoms defining a clinical condition (Nemiah, Freyberger, & Sifneos, 1976; Sifneos, 1973). Sufferers of alexithymia could not identify their emotions (often confusing them with other physical sensations), could not describe their emotions, had a concrete thinking style oriented to external rather than internal events and lacked a rich internal world of fantasy and imagination. Alexithymia is now understood to exist more as a continuum than a clinical condition (see e.g., Parker, 2004). It is frequently measured by the Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994a), which covers three of the four aspects of alexithymia: difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and an externally-oriented thinking style (EOT). There are obvious theoretical links between these three facets of alexithymia and Strategic EI. Verbally expressing emotions (DDF) relies on a basic understanding of emotions. Distinguishing when emotions (rather than physical sensations) occur (DIF), and devoting thought to emotional phenomena (EOT) are pre-requisites for both managing and understanding emotional situations. Warwick and Nettelbeck (2004) found the MSCEIT showed small relationships with two of alexithymia's facets (difficulty in identifying and describing feelings), and this relationship was stronger for Facilitation and Understanding branches. Barchard and Hakstian (2004) also found a relationship between these two aspects of alexithymia and performance-based EI.

Academic achievement. Using academic achievement as a criterion to assess the validity of EI is somewhat controversial as the justification for such a link is quite indirect (Landy, 2005). Contextual social factors relating to EI might predict increments in performance that cognitive ability does not capture (e.g., getting along with classmates or instructors, negotiating study leave with employers, obtaining help with homework from parents, siblings, or housemates). Nevertheless, much prior research linking EI with academic achievement has been conducted (Barchard, 2003; Newsome et al., 2000; Parker, Summerfeldt, Hogan, & Majeski, 2004; Petrides, Frederickson, & Furnham,

2004; van der Zee, Thijs, & Schakel, 2002; Woitaszewski & Aalsma, 2004). However, only a few of these examined *performance-based* EI in detail. Barchard found that MSCEIT Understanding scores (but not Perception, Facilitation, or Management) predicted academic performance at University, but that this was not significant after personality and intelligence were accounted for. Woitaszewski and Aalsma did not find a relationship between the MEIS and academic achievement even before further variables were accounted for (although their sample comprised gifted students only, and thus may have had restriction of range effects). Van Rooy and Viswesvaran's meta-analysis (2004) reported a trivial average correlation of .09 between EI and academic performance (compared to a correlation of .22 between EI and workplace performance). However, the meta-analysis aggregated self-report and performance-based indices of EI and as such may underestimate the prediction of academic performance by performance-based EI (assuming that the more personality-like self-report measures are *less* predictive of academic performance than the more intelligence-like performance-based measures of EI).

If the STEU and STEM do capture aspects of contextual social factors, then they ought to predict academic achievement in undergraduates over and above the effects of cognitive ability and personality. Chapters 3 and 6 examine whether these newly constructed tests do predict any additional variation in academic achievement at university once intelligence and personality are accounted for.

7. Goals and Aims

7.1 Test Construction

The primary aim of this research is to establish whether it is possible to measure performance-based EI using appraisal theory and the SJT paradigm to create tests, and to obtain meaningful systems of scoring that do not rely solely on group judgment. The validation of the STEU (created and scored by appraisal theory) and STEM (developed according to the SJT technique) involves their relation to other measures of EI, to intelligence, to lower levels of alexithymia, emotional distress and loneliness, to higher levels of life satisfaction and academic achievement and lifestyle variables such

as income, job status, social contact, and substance use. In addition, *incremental validity* of the STEU and STEM is examined by controlling for personality and intelligence in the relationship between EI and criterion measures such as stress, loneliness, life satisfaction, and academic achievement.

7.2 Examination of "Method Effects" in Assessing Performance-Based EI

The possibility of two "method effects" in the relationship between different tests of EI, and between EI and intelligence, will be examined: (1) the ratings-based format used in all non-Understanding branch MEIS and MSCEIT tests; and (2) consensus scoring. It is hypothesised that ratings-based item format and consensus scoring used in EI tests have resulted in overestimation of relationships between EI tests and underestimation of relationships between EI tests and intelligence. Ratings-based and multiple-choice formats are directly compared in Chapter 3, with the expectation that EI test scored will relate more strongly to each other when their response formats are the same, and more strongly to intelligence when their response format is multiple-choice rather than ratings-based. In Chapter 5, the distinction between EI and Gc is evaluated using multiple-choice tests of EI to control for format method effect, and by comparing two solutions: (1) when Gf and Gc tests are scored conventionally (i.e., standards-scored) but EI tests are consensus-scored (as is generally the case in analyses of the MSCEIT), and (2) when all measures (including Gf and Gc) are consensus-scored. It is expected that Strategic EI tests will be strongly related to Gc, and that this relationship will increase when Gc tests are also consensus-scored.

7.3 Validity of Emotional Intelligence

7.3.1 Distinction of Performance-Based Strategic EI from Gf and Gc

It is predicted that performance-based EI will relate more strongly to Gc than to Gf for two reasons. Firstly, the verbal load of text-based tests may function as a method factor. Secondly, the constructs of emotional understanding and management are assessing *knowledge* of the situational antecedents to emotions, as well as the most effective methods for ameliorating negative emotions. Whether the relationship between EI and Gc is due to a verbal ability contamination in the EI tests, or is a meaningful relationship, is also examined. Chapter 5 controls for scoring and response-format

method factors and Chapter 6 examines the relationship between verbal measures of Strategic EI and Gc after reading comprehension is accounted for, and also the relationship between Strategic EI (where items are text-based) and emotion recognition ability (where items are *not* text-based but nonverbal).

7.3.2 Strategic EI and Prediction of Emotion-Related Criteria

To evaluate the validity of the STEU and STEM as measures of performance-based EI and of the validity of performance-based EI as a construct, the STEU and STEM's relation to emotion-related criterion variables is examined. Chapter 3 compares the STEU and STEM to alexithymia, life satisfaction, and mental health in an undergraduate sample. Chapter 3 also examines whether these measures can incrementally predict academic achievement once intelligence and personality are accounted for. Chapter 4 looks at the power of the STEU and STEM to predict states of depression, anxiety and stress in a sample of volunteers drawn from central Sydney. Chapter 5 re-examines the STEU and STEM's relationship to alexithymia, state stress and life satisfaction but controls for personality, fluid intelligence, and crystallized intelligence in these relationships. Similarly, Chapter 6 re-examines the prediction of academic achievement but controls more extensively for intelligence and personality than Chapter 3.

7.3.3 Validity of Self-Report EI

A brief examination of the overlap between self-report EI and the 30 faceted NEO-PI-R model of personality is also undertaken in Chapter 5. After examining the structure of the Schutte Self Report Scale (SSRI; Schutte et al., 1998), the independence of each of the SSRI's components from the NEO-PI-R personality model is assessed.

8. Concluding Comments for Chapter 1

In asking "what is EI good for?" answers should depend on what intelligence and personality are already known to be good for – EI's prediction of these emotion-related criteria is only potentially useful if they cannot already be predicted by other means. Landy (2005) and Matthews et al. (2002) point to the paramount importance of assessing incremental validity over and above both intelligence

and personality variables (and over fairly exhaustive measures of each, at that). Few studies have controlled for both broad measures of intelligence and broad measures of personality (Bastian et al. [2005] provide an important exception, with the MSCEIT incrementally predicting anxiety but not life satisfaction, problem solving, coping style or academic achievement). Both intelligence and the five factor model of personality are controlled for in studies in Chapters 3, 5, and 6 when examining construct validity. In addition, Chapters 5 and 6 control for both Gf and Gc.

Chapter 2

Test Development of the Situational Test of Emotional Understanding (STEU) and Situational Test of Emotion Management (STEM)

This chapter describes the development of two performance-based EI tests: the Situational Test of Emotional Understanding (STEU; outlined in Part I of this chapter) and the Situational Test of Emotion Management (STEM; outlined in Part II of this chapter). The STEU was developed according to an empirically-derived model of the structure of emotions (Roseman, 2001; Roseman et al., 1996), and the STEM according to Situational Judgment Test (SJT) methodology. Although tests of Understanding and Management already exist in current EI batteries (i.e., the MEIS and MSCEIT; see Mayer et al., 1997, 2002a), there are at least two reasons why developing new ones using different methods might be beneficial: (1) the need for standards-scored tests, and (2) using different scoring and response-format methodologies to ensure that characteristics of EI tests are characteristics of the EI construct rather than the measurement method.

Standards-based scoring of EI tests is important for test validity theoretically, empirically, and practically. Theoretically, it is important for construct validity to define why an answer is correct (i.e., rather than a black box of scoring algorithm, it should be possible to know what types of knowledge or cognitive skill are used to answer the question). Without knowing what people are doing when they answer an item, it is virtually impossible to use a cognitive psychology approach as a method for construct refinement. Empirically, test score distribution shape may be skewed and leptokurtic under consensus scoring (MacCann et al., 2004). Practically, if tests are used in high stakes decision making, legal defensibility of why certain answers are wrong may become an important issue.

With only one existing measure of performance-based EI, it is possible that unusual aspects of the methodology (such as the ratings-based judgment response format, or consensus-type scoring) may be responsible for current substantive findings. Alternative tests that measure the same

constructs with different methods are needed to ensure this is not the case. This chapter outlines the development of two alternative tests. Part I describes item development for the STEU based on Roseman's (2001) appraisal theory of the structure of emotions. Part II describes the development of items and scoring systems for the STEM in a series of four studies based on the SJT approach to test construction: (1) item generation from responses to a semi-structured interview; (2) response-option generation from free responses to options (including an examination of whether personality affects these responses to situations); (3) construction and validation of an expert scoring key; and (4) construction of a standards-based scoring system

Part I of Chapter 2

Development of the Situational Test of Emotional Understanding (STEU)

1. Appraisal Theories of Emotion: A Foundation for Standards-Based Scoring of Emotional Understanding Tests

Numerous studies of retrospective reports of emotional situations have been conducted in appraisal theory research, where analyses aimed to discover lawful rules that linked particular antecedents to particular emotions (e.g., Roseman, 1991; Roseman et al., 1996; Roseman, Spindel, & Jose, 1990; Roseman et al., 1994; Scherer, 1993, 1999a; Smith & Ellsworth, 1985). In general, appraisal theories of emotion hold that appraisal (a combination of perception and judgment perhaps best thought of as apperception) of the surrounding circumstances is responsible for both the elicitation of emotion, and for the differentiation of emotion (see e.g., Frijda & Zeelenberg, 2001; Schorr, 2001). That is, appraising a situation as representing some combination of abstract dimensions or a single "core relational theme" generates a particular emotion (e.g., Lazarus, 1991; Smith & Ellsworth, 1985; Roseman, 1984; Roseman, 2001; Scherer, 2001).

Roseman's (2001) model of the emotions system was selected as the basis for construction of a standards-scored test of emotional understanding for two reasons: (1) an appropriate level of detail for easy translation from theory to item phrasing, and (2) empirical evidence. Roseman's theory has few enough appraisal dimensions so that 1-2 sentences are adequate to describe all dimensions relating to a particular emotion. Two other prominent appraisal theories were rejected for having too much and too little detail respectively.[5] Roseman's (2001) model consists of seven appraisal dimensions, outlined in Table 2.1. Each dimension has two or three possible values easily translated into everyday speech (e.g., a motive consistent situational state can be phrased as "something you want happens").

[5] Scherer's (1999b; 1988) sequential check theory was rejected for being too detailed. Sequential check theory meticulously relates up to 32 appraisal dimensions to each emotion, via an algorithm too complex to describe without the item becoming a measure of reading comprehension rather than understanding the causes of emotions. Lazarus' (1999) core relational themes were rejected for too little detail, as there was a one-to-one correspondence between appraisal dimension and emotion (e.g., anger as "a demeaning offence against me and mine"; Lazarus, 1999, p. 64). It was thought that this one-to-one correspondence might make items far too easy for between-subjects variability to emerge in individual differences testing. Note that the selection of theory was based more on the utility of the theory for the current purposes, rather than a detailed analysis of which appraisal theories were the most valid: i.e., Roseman's theory is more *useful* for present purposes, rather than more *valid* than other theories.

Table 2.1

Explanation of Each Appraisal Dimension and its Possible Values in Roseman's (2001) Model for the Structure of the Emotions

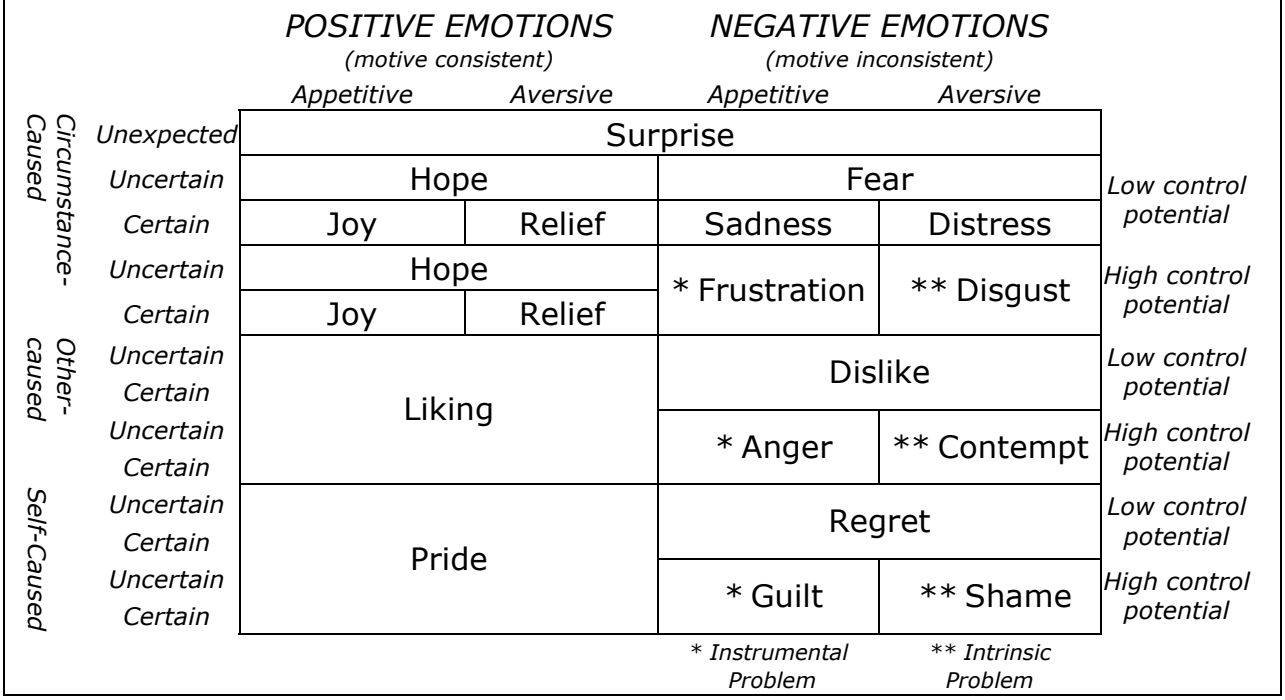
Dimension and Definition	*Appraisal Values*
1. Unexpectedness: whether the event violates one's expectations	➢ Expected ➢ Unexpected
2. Situational state: whether an event is wanted or unwanted by the person	➢ Motive-inconsistent ➢ Motive-consistent
3. Motivational state: whether the event relates to a desire to get less or something punishing or a desire to get more of something rewarding	➢ Aversive (minimise punishment ➢ Appetitive (maximise reward)
4. Probability: whether the occurrence of motive-relevant aspects of the event is merely possible, or is definite	➢ Uncertain ➢ Certain
5. Agency: who or what caused the motive-relevant event	➢ Circumstance caused ➢ Other person caused ➢ Self caused
6. Control potential: whether there is nothing one can do, or something one can do about the motive-relevant aspects of the event.	➢ Low control potential ➢ High control potential
7. Problem type: whether a motive-inconsistent event is unwanted because it blocks attainment of a goal or unwanted because of some inherent characteristic	➢ Instrumental problem (blocks attainment) ➢ Intrinsic problem (inherent characteristic)

Roseman's (2001) theory uniquely defines 17 emotions by combining the seven appraisal dimensions as outlined in Figure 2.1. For example, relief is caused by appraisals of circumstance-cause, certainty, motive consistency, and aversive stimuli. This structure of the emotions has been empirically tested and refined using three different methodologies: (1) vignette research (Roseman, 1984; Roseman, 1991); (2) retrospective studies (Roseman et al., 1990, 1994, 1996); and (3) experimental research (Roseman & Evdokas, 2004). Cross-cultural comparisons have also been conducted (Roseman, Dhawan, Rettek, Naidu, & Thapa, 1995). In vignette research, participants are presented with different versions of the same story where appraisal dimensions have been manipulated and asked to rate the emotions experienced by the story's protagonist (e.g., in some versions, the event may be motive consistent, in other versions, the event is motive inconsistent). Retrospective research involves participants recalling events they have experienced that caused them

to feel a particular emotion. After recalling these experiences, they then answer a number of questions about how they appraised the event in terms of the proposed dimensions. Experimental research manipulates participants' expectations or interpretations of what is happening (i.e., their appraisals) and then evaluates whether different appraisals do result in the proposed emotions being generated.

Figure 2.1

Roseman's Structure of the Emotions (from Roseman, 2001, pp. 70-71)



2. Using Roseman's Model to Generate Items for the STEU

In the current thesis, items were generated by translating this model's appraisal dimensions into lay-language. For example, relief-generating appraisals of circumstance-cause, certainty, motive consistency, and aversive stimuli were translated into "An unwanted situation becomes less likely or stops altogether". Participants are then asked how the person involved is most likely to feel and given five possible emotions including the correct response (relief, in this case). The phrase "most likely to feel" was deliberately chosen (rather than "would feel" or "the strongest emotion they would feel"),

as the empirical evidence for this theory is "proof" in a probabilistic sense rather than a contingent sense (i.e., statistically significant relationships between appraisal dimension and emotion).

A similar process was conducted for all the other emotions in the model (except for guilt, shame, and disgust), resulting in 14 items phrased in this abstract sense.[6] For each of these 14 items, analogues representing workplace and personal-life contexts were created. For example, "An unwanted situation becomes less likely or stops altogether" had the workplace analogue "A supervisor who is unpleasant to work for leaves Alfonso's work" and the personal-life analogue "An irritating neighbour of Eve's moves to another state." Thus the STEU was constructed be a 42 item test with 14 context-less items, 14 items relating to personal life, and 14 items with a workplace context. These items are reproduced in Appendix 2.1. This division of the STEU into 14 possible emotions and 3 possible contexts (no context, work context, and personal-life context) is used to divide the STEU into subsets in Chapter 3 and Chapter 4. Chapter 3 considers the three contexts as subsets of the STEU. Chapter 4 divides the STEU into 14 item parcels by emotion.

3. Final Description of the STEU

The STEU thus assesses Understanding in terms of understanding the emotions that are occurring in different situations. The STEU is scored according to empirically verifiable facts (i.e., an empirically-derived theory on why certain situations result in certain emotions), rather than group judgment. This is an important advantage of the STEU over existing measures of EI, as it allows verification of the correct answer. Evidence for the validity of the STEU as a measure of Emotional Understanding is collected in Chapters 3, 4, 5 and 6.

---

[6] These were not included, as it was felt that specific situations for guilt and shame might be too specific to individuals' interpretations and personality, and situations involving disgust would be too obvious and easy, resulting in a lack of variance.

Part II of Chapter 2

Development of the Situational Test of Emotion Management (STEM)

The Situational Judgment Test (SJT) methodology used by many I/O psychologists to assess noncognitive qualities was here used to assess the management of emotional situations (Management). Despite early claims to the contrary (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001), there is growing consensus that SJTs are best conceptualised as a methodology rather than as a theoretical construct (Hanson, Horgen, & Borman, 1998; McDaniel & Nguyen, 2001; McDaniel & Whetzel, 2005). The breadth of constructs now measured by SJTs supports this conclusion (Funke & Schuler, 1998; Hedge, Bruskiewicz, Borman, Hanson, & Logan, 2000; Hedlund et al., 2003; Kline, 1994, 1996; Motowidlo et al., 1990; Sternberg et al., 2001; Wagner & Sternberg, 1987; Wagner, Sujan, Sujan, Rashotte, & Sternberg, 1999). Hanson et al. (1998) argue that the methodology is particularly well-suited to interpersonal skills, and constructs involving judgment and decision making in complex situations. Thus, it seems both reasonable and worthwhile to use SJTs as a methodology for creating an emotion management measure.

McDaniel and Nguyen (2001) review the methodology used to construct SJTs, summarising the process into the following three steps: (1) *item generation*: situations are collected from a relevant sample and used to create item stems; (2) *response option generation*: item stems are presented to a second relevant sample who generate responses to the situations which are used to create response options; (3) *expert scoring*: this final set of items is presented to a group of experts who determine what the best response options are. To construct the STEM, three studies corresponding to these three steps are presented below: Study 1 generates items, Study 2 generates response options, and Study 3 generates a scoring key based on expert judgment.

In Study 1, semi-structured interviews based on the elicitation of emotional situations were held with undergraduate students and community volunteers. Responses were summarised via qualitative content analyses to provide item stems. In Study 2, a second sample of undergraduates

wrote down their actual and ideal responses to these item stems, and response options were generated from qualitative analyses of these responses. Study 2 also compared responses generated by asking for the "best" behaviour to responses generated by asking "what would you do" in terms of response variability and relationships with personality dimensions. Study 3 involves a group of counsellors, EI researchers and life coaches evaluating the response options to determine expert scoring weights. In addition to expert judgment, a scoring code based on standards was derived from psychological research. More specifically, literature from the coping field was summarised into a scoring code that was used to score the items of the STEM. The development of this standards-based approach to scoring in described in Study 4. These four studies are described below.

1. Study 1: Generation of Situations

1.1 Introduction to Study 1

The purpose of this study was to collect descriptions of situations where emotions occurred and needed to be managed, so that such descriptions could be summarised into items for an SJT assessing Emotion Management (i.e., the STEM). Descriptions of emotional situations were collected via semi-structured interviews, and then distilled into short excerpts to form items on the STEM. The study's design was based on three steps outlined below: (1) defining what constitutes a relevant sample when the construct of interest is Emotion Management; (2) obtaining relevant and broad content coverage of the Management construct, and (3) summarising content into items with maximum efficiency and minimum loss of information.

1.1.1 Step 1: Selection of a Relevant Sample

Virtually anyone would constitute a relevant sample for collecting emotional situations. Anyone living in a society has relatively extensive experience with emotional events, and has had to manage many of them during the course of ordinary life (except perhaps alexithymics or those with William's syndrome or similar deficits of emotion). However, test bias issues in situational questions (where different types of situations will be more familiar or more meaningful to different age groups, cultural groups, genders, and socio-economic groups) attest to the importance of sample diversity.

Thus, community volunteers were used in addition to a sample of undergraduate psychology students. In addition, different *types* of situations were elicited in the semi-structured interview to ensure a broad content coverage of types of situations (see Section 1.1.2 below for a description of these different types of situations).

1.1.2 Step 2: Content Elicitation of a Broad Range of Situations

To ensure a wide range of incidents were covered, content elicitation was structured according to both domain of life, and type of emotion. Scherer and Tannenbaum's (1986) qualitative analysis of emotional situations found that the most frequently coded content characteristics were work, family, friends, and the behaviour of strangers. This division of incidents into the domain of life (i.e., work, family, friends) was used in the current study. Work and family categories were also conceptually expanded into several underlying categories (e.g., work consists of boss, peers and subordinates/employees; family of parents, partner, children, siblings). Thus participants were asked to describe several different types of situations where emotions might occur.

In eliciting different emotions, an important issue was *which emotion types* to select, given that different theorists disagree as to the number and type of emotions that exist (Ekman, Friesen, & Ellsworth, 1972; Izard, 1971; Izard, 1993; Oatley & Johnson-Laird, 1987; Plutchik, 1997; Tomkins, 1968). Rather than attempt to resolve the thorny issue of whether, and which, emotions are "basic" (see Ekman, 1992a; Ekman, 1992b; Izard, 1992; Ortony & Turner, 1990; Turner & Ortony, 1992 for a sample of that debate), this study takes the view that the six emotions of joy, sadness, fear, anger, surprise, and disgust (the universal facial expressions named by Ekman, Friesen, and Ellsworth in 1972) can safely be used as a representative sample of discrete emotions. Such emotions are contained in most theories, and have some evidence of cross-cultural expressive similarity, and neurological correlates (Ekman et al., 1987; Panksepp, 1992). This list is not assumed to be exhaustive or definitively correct, as it excludes the self-conscious emotions and does not differentiate between positively valanced emotions. However, in the interests of efficiency it would appear a reasonable starting point for the collection of emotional situations.

1.1.3 Step 3: Translation of Responses into Item Stems

1.1.3.1 Selection of Situations

Motowidlo et al. (1997) propose that SJTs might be developed by categorising situations into broad content areas, and then selecting specific situations representative or prototypical to these content areas. For example, "handling difficult customers" might be one content area, and all specific instances of this could be reviewed and a prototypical situation representing "handling difficult customers" could be created from common elements or representative scripts. A variation of this approach was used in the current study. Situations were classified as: (1) general themes (e.g., Fight, Celebration); (2) specific content areas (e.g., Shopping for Clothes with Friend); and (3) as representing one or more emotions (Joy, Sadness, Fear, Anger, Disgust, Surprise). Several general themes associated with each emotion were selected for item creation. For each theme, several specific examples were translated into items based on their common elements.

1.1.3.2 Summarising Situations into Items

Once collected, the scripts of situations were transformed into item stems by summarising content into a short description of key elements of the situation. A 2-sentence maximum length for items was set for two primary reasons: (1) evidence that use of long verbatim scripts results in low test reliability; and (2) long text-based items may result in the measurement of reading ability or verbal ability rather than EI (as may have been the case for social intelligence; see Landy, in press).

(1) Low reliability of verbatim scripts. Mayer and Geher (1996) used verbatim scripts describing emotional situations as item stems for an instrument assessing emotion recognition in situations, and found very low reliabilities for this instrument (the Emotional Accuracy Research Scale, or EARS). Although low reliabilities may have been due to the forced-choice response format used, or the fact that SJTs tend to be multi-dimensional and hence have low internal consistencies (see Legree, 1995; and Chan & Schmitt, 2002, respectively), low reliabilities may equally have been due to the incidental detail involved in verbatim scripts. To avoid irrelevant details that may affect

participants' responses to the test items, it was decided to use short, clear and explicit descriptions of the key details of emotional situations (rather than verbatim scripts) in creating STEM items.

(2) Text-based items and construct contamination from reading ability. The history of social intelligence measurement is replete with difficulties in establishing the social intelligence construct as separate from other forms of intelligence (Kihlstrom & Cantor, 2000). Landy's summaries (2005; in press) of the history of social intelligence suggest particular overlap between verbal text-based measures of social intelligence and verbal ability. Although SJT items are ordinarily long enough such that items are rich in detail and correspondingly ecologically valid, such length also increases the reading age required to answer such items. The two-sentence maximum rule trades this rich level of detail for a lower threshold level of verbal skill, in line with concerns that such items should not be proxies for verbal ability.

1.1.4 Aims and Broad Overview of this Study

The broad aim of this study is to collect descriptions of situations where people experience different types of emotions. By classifying each description of a situation as representing an emotion, and a broad general theme, several broad categories of situation can be included for each of several emotions. This allows broad content coverage over both situation type and emotion type that is specifiable, such that each item generated represents a particular type of event that is associated with a particular emotion.

1.2 Method

1.2.1 Participants

50 participants (30 female) completed the half-hour interview. 31 of these participants were undergraduate psychology students at the University and Sydney (7 of whom were mature age students) and 19 were community members recruited through Rotary clubs, churches, Parents and Citizens committees, and Police and Citizens Youth Clubs. Age ranged from 17 to 59 years with a mean of 31 (median = 22; *SD* = 14.8). 15 of the participants were married, and 6 were non-native English speakers.

1.2.2 Materials: Semi-Structured Interview

The format and phrasing of the semi-structured interview was similar to previous interview research into emotional episodes (e.g., Fitness, 2000; Fitness & Fletcher, 1993; Scherer & Tannenbaum, 1986). Participants were asked to "think back over the last couple of weeks and recall a time when you felt some kind of emotion. Think about what was happening to you at this point and describe the situation in your own words". Respondents were prompted for more detailed description if necessary (e.g., Where were you? Who were you with? What was happening? What emotion were you feeling? Can you give me more detail about that?).

There were two alternate forms of the interview. The first (Form A) was based on six emotions – interviewees were asked to describe a separate situation where they had felt joy, sadness, anger, fear, surprise, and disgust respectively. Form B asked for situations based on the identity of the other person(s) involved. Participants were asked to describe situations involving their family, partner, children, siblings, parents, friends, work acquaintances, co-workers, boss, and employees, where these categories were relevant. In addition, the first question on both forms was general, in that it asked for a description of a situation where the participant experienced "some kind of emotion". Thus, each interviewee described from three to eleven episodes, for a total of 290 episodes. Interviews took approximately 20 minutes each.

1.2.3 Procedure

All interviews were conducted by the author. Interviews were held face-to-face for all undergraduate students and for community members where possible (several interviews were held over the phone). Interviews were tape-recorded and transcribed, and the transcriptions entered into the Non-Numerical Unstructured Data Indexing Searching and Theorising Version 6 (NUDIST or N6) software package for coding and analysis.

1.2.4 Coding

In N6, coding operates by the creation of a number of categories called *nodes*. Each unit of information (or case) can be coded as representative of one or more nodes. In the current study, each

participant's response to each question was considered a separate case. This meant there were 290 separate cases, each acting as a single unit of text that could be coded at various nodes. Certain classes of nodes were set up a priori, before examination of the data. These included nodes representing *basic emotions* (Joy, Sadness, Fear, Anger, Disgust, and Surprise); *domain of life* (Family, Work, and Friends); and *emotional intensity* (High Intensity, Medium Intensity, and Low Intensity).

Two other classes of nodes were created after analysing commonalities among the described situations: *general themes* and *specific themes*. General themes were a broad description of the item content. Once at least three cases appeared to represent a certain type of event, a node was created for this type and then subsequent cases representing this event type were coded at this node. After coding, 18 general theme nodes coded five or more cases and these were retained for analysis (e.g., Fight, Goal Striving Impeded, Secrecy or Lies; see Table 2.2 for the full list of general themes). A brief content label was created for each case, and these became nodes for *specific themes* (e.g., Flatmates Won't Pay for Shared Household Items, Visit Mother in Nursing Home). Some cases were so similar that they could be coded under the same *specific theme* (in the end, there were 168 specific themes for the 290 cases). An example of text coded at the specific theme node "Public criticism by boss", and also coded as representing anger and high emotional intensity is given in the paragraph below for the purpose of illustration:

> *Once she gave me a dressing-down in front of a customer, which no manager should do. And she accused me of being idle, standing there having my arms folded but that wasn't the case. I mean we had to mark out territories because she's very territorial like that because of shop lifting and if there's no customers there there's only so much folding you can do. I think she had a bad day or something and she took it out on me, and she's taken it out on others so I wasn't alone in that but that really did piss me off.*

1.2.5 Planned Analyses

The text generated in response to the semi-structured interviews was analysed in several steps to get to the final set of items. These steps are outlined below.

Analysis 1: Reliability Check. The inter-coder reliability of coding at each of the 6 emotions nodes, 18 general theme nodes, emotional intensity node and domain of life node is calculated to ensure node categories are reliable.

Analysis 2: Check on Causes of Emotion: Situation or Person-Caused? Multi-level modelling determines whether significant amounts of the variance in the 6 emotions nodes is due to person-level factors rather than situation-level factors. This is to ensure that the emotions represented in the collected situations are due to situational details rather than the emotionality of the people reporting them.

Analysis 3: Emotion occurrence. The data are described in terms of how many situations were collected for each emotion, and how many of these situations involved more than one emotion.

Analysis 4: General Themes Linked to Emotions. The data are described in which general themes tended to occur for situations representing each of the six emotions.

Analysis 5: Creation of Items from General Themes, Specific Themes, and Reference to Text. A full description and illustrative example shows how 1-2 sentence items are created from specific themes and reference to the verbatim text of the interview responses.

1.3 Results and Discussion

1.3.1 Reliability of Coding

100 of the 290 cases were re-coded by a research assistant in order to estimate the reliability of the coding. Because of the relative scarcity of "hits" (cases coded) compared to "misses" (cases *not* coded) raw level of agreement between coders can be fairly high simply by both coders not coding much. For this reason, additional parameters of reliability were estimated for different categories of nodes, in line with the conceptual measurement properties of those nodes. Polychoric correlation coefficients between coder 1 and coder 2's rating were calculated where the underlying construct was thought to be continuous (emotional intensity, general theme [where incidents might resemble the theme to a certain degree], and emotion [where incidents might resemble a particular emotion to a certain degree]), and Cohen's Kappa was calculated for categorical data (the domain of life that the

incident occurred in). Even here, the very high rate of misses meant that reliability parameters given may be over-estimates. Table 2.2 gives the raw rate of agreement, and the reliability coefficient for all coding categories.

Table 2.2

Descriptions of Coding Categories, and Their Raw Level of Agreement and Reliability for Coder 1 and Coder 2's Coding

Category	*Description of Category*	p_0	*Reliability*
General Themes			
Achievement	Succeeding at something or successfully reaching a goal/completing a task.	.95	.93
Put Down or Rudeness	Someone putting you down, insulting you, or behaving rudely towards you (*originally two categories: rudeness and put down*).	.95	.97
Unfairness	Perception by the interviewee that events were not fair, just or equal.	.88	.81
Role Change	A change in duties, or change in primary daily activities (e.g., Promotion, demotion).	.95	.84
Help Others	Helping other people (e.g., giving, teaching, comforting).	.93	.69
Lack of Time	The perception that there is not enough time to do all activities necessary in one's life.	.95	.81
Celebration	A gathering to celebrate something (e.g., birthday, Christmas, anniversary etc).	.95	.88
Morally Wrong or Taboo	The interviewee disapproves of events, behaviours or thoughts for moral or ethical reasons.	.96	.90
Anticipation	Emotions generated prior to expected or imagined event.	.90	.77
Goal Striving Impeded	Obstacles in the path of getting to the goal, or getting what you want -- something goes wrong that interferes with what you want to happen.	.98	.97
Affiliation	Perception by the interviewee that a relationship, closeness or connection to others is felt, or is increasing.	.88	1.00
Parting	Parting company from someone you are close to (distinguished from terminating a romantic relationship, which was a separate category excluded for infrequency).	.94	.54
Health of Others	A change in the health or friends or relatives.	.97	.88
Losing Closeness or Friendship	The perception that a connection to others in losing intensity or meaning, or is ending altogether.	.93	.75
Loneliness	Feeling isolated from others, lacking company when one wants it.	.98	.86
Illness	One's own illness.	.93	.61
Disorganisation About Meeting	Difficulty in meeting others at pre-arranged time or place.	.98	1.00
Fight	Where a fight, conflict or argument occurred between two or more people (where both or all participated).	.94	.91

Table 2.2 (continued)

Category	*Description of Category*	p_0	*Reliability*
Emotions			
Joy	One of primary emotions felt was joy.	.91	.95
Sadness	One of primary emotions felt was sadness	.85	.81
Fear	One of primary emotions felt was fear.	.96	.98
Anger	One of primary emotions felt was anger.	.95	.99
Disgust	One of primary emotions felt was disgust.	.98	1.00
Surprise	One of primary emotions felt was surprise.	.94	.94
Emotional Intensity	Intensity of the emotion: high, medium or low	.53	.81
Domain	Domain of the situation, rated as family, work, romantic attachment, strangers, no other people involved, or flatmate	.86	.84

Note. Reliability estimates are polychoric correlations, except for the category Domain where the reliability estimate is Cohen's Kappa.

All reliability estimates were reasonable except for Parting (with a reliability estimate of .54) and Emotional Intensity (where the raw level of agreement was .53). Data was re-examined to see why this might be so. It transpired that the second coder had included dyadic or romantic parting (relationships breaking up) within Parting whereas coder 1 had not. The coders also disagreed on Medium Intensity versus High Intensity – 27 cases were coded High Intensity by coder 1 and Medium Intensity by coder 2 (see the contingency tables in Appendix 2.2). Because of this, the categories Medium Intensity and High Intensity were combined. The raw agreement for this demarcation (High/Medium Intensity and Low Intensity) was .87, and the polychoric correlation between raters was .83.

1.3.2 Situation-level and Person-level Effects

The purpose of this study is to collect situations that represent different emotions, so the emotion's occurrence should be *situation*-caused more than the person-caused (e.g., anger should be caused by the unfairness in the situation more so than by being a generally angry person). This data is structured such that situations constitute the cases, which are then grouped within the individuals who report them. Such a structure allows testing of the proportions of variance in emotion occurrence (i.e., whether or not each emotion occurs in a given situation) due to the situation, and due to the person. A data file was created with a field for each emotion of joy, sadness, fear, anger, disgust and surprise,

each case coded as "1" for the emotion's presence, and "0" for the emotion's absence, and marked for the individual person that it came from. Variance components models were then computed to determine the amount of variance at the *person level* and at the *situation level* responsible for the presence of the emotion. The proportion of variance in emotion occurrence due to situation and person factors was calculated and is shown in Table 2.3 for Joy, Anger, Sadness, Fear, Disgust, and Surprise. In all cases, the vast majority of the variance (greater than 90% for all but joy) in emotion occurrence was due to situation-level rather than person-level factors. However, the 11.6% of variance in joy due to person-level factors was significant, meaning that joy is caused both by joyful situations and by being a generally joyful person.[7] Given that the majority of variance is due to situation level rather than person level factors, further analyses will ignore the person level and concentrate only on the situation level in further analyses.

[7] Note that the trivially small amount of variance in disgust elicitations (at either level) has resulted in a non-meaningful negative standardized value.

Table 2.3

Variance Components Models Predicting the Proportion of Variance in Emotion Presence/Absence Attributable to Person-Level and Situation-Level Factors for Joy, Sadness, Fear, Anger, Disgust, and Surprise

	Presence/Absence of Emotion Person Level				*Presence/Absence of Emotion Situation Level*			
		Estimate	*SE*	*Z-value*		*Estimate*	*SE*	*Z-value*
Joy	11.6%	.015	.007	2.08*	88.4%	.114	.010	10.99**
Sadness	7.1%	.008	.006	1.52	92.9%	.105	.010	11.00**
Fear	4.6%	.004	.004	1.04	95.4%	.083	.008	11.02**
Anger	6.3%	.011	.008	1.36	93.7%	.164	.015	11.01**
Disgust	0.0%	.000	.000	-1.08	100.0%	.010	.001	10.84**
Surprise	0.0%	.000	.002	0.14	100.0%	.059	.005	11.04**

*$p < .05$, **$p < .01$

1.3.3 Frequencies of Coding

Table 2.4 shows the number of situations representative of each emotion, including situations representing more than one emotion, and the relative emotional intensity for each category of emotion. Three things are immediately apparent. Firstly, many more anger-related situations were elicited (105 of 290 cases), than situations related to other emotions (particularly disgust and surprise). Secondly, Joy is felt less intensely than the other emotions (significantly so for Anger, Disgust, and Surprise, at $p < .05$).[8] Thirdly, rather than occurring in isolation, Surprise frequently combined with Joy or Sadness. As an unvalenced emotion, Surprise seems to behave differently from other emotions, and it was decided not to include Surprise items in the STEM. Joy management items

[8] χ^2 goodness of fit indices were calculated using the expected values based on the relative frequencies of the emotions.

were also not included, primarily due the significant person-level variation in Joy elicitation, but also because of the lower intensity of Joy compared to other emotions.

Table 2.4

Frequency of Emotion Co-occurrence (Number of Cases on Diagonal), and Emotion Intensity

Emotion	*Emotion*						*High/Medium*		*Low*	
	Joy	*Sadness*	*Fear*	*Anger*	*Disgust*	*Surprise*	*N*	*%*	*N*	*%*
Joy	78						55	71	18	23
Sadness	2	60					46	77	12	20
Fear	6	5	55				41	75	11	20
Anger	1	11	11	105			79	75	10	10
Disgust	1	3	0	8	22		16	73	2	9
Surprise	19	7	1	4	1	34	26	76	5	15

Note. It was not possible to meaningfully code intensity for all cases (thus frequencies for intensity do not add up to total frequency).

Table 2.5 shows the number of cases representing each emotion at each broad general theme (i.e., the link between general theme and emotion type). Themes associated with joy were (unsurprisingly) Achievement, Affiliation, and Celebration. Sadness was associated with Health of Friend/Relative, Declining Friendship/Closeness, Illness, Parting, and Loneliness/Isolation. Fear was associated with Anticipation, Health of Friend/Relative, Role Change, and Illness. Anger was associated with Fight, Putdown/Rudeness, Goal Striving Impeded, Unfairness/Injustice, Disorganisation about Meeting, and Lack of Time. The association of anger with unfairness has been examined before by Mikula, Scherer, and Athenstaedt (1998) who found that the perception of injustice was related to all negative emotions but most strongly to anger.

To include content from at least two general themes for three major emotions (sadness, anger, and fear), the following themes were selected as the framework for deriving item content: (1) Fight; (2) Anticipation; (3) Put Down/Rudeness; (4) Unfairness; (5) Role Change; (6) Parting, Loneliness, and Losing Friendships/Closeness (combined to form the category Isolation based on their conceptual

similarity); (7) Disorganisation About Meeting (conceptually expanded to also include other communication issues); (8) Goal Striving Impeded; (9) Morally Wrong or Taboo; and (10) Health of Friends/Family, and Illness (combined to form the category Health Concerns). For each of these general themes, several related specific themes were selected. The list of specific themes (organised by emotion and intensity level) is given in Appendix 2.3.

Table 2.5

Number of General Themes Coded at Each Emotion

General Themes	*Total*		*By Emotion*											
			Joy		*Sadness*		*Fear*		*Anger*		*Disgust*		*Surprise*	
	N cases	*N spec themes*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
Achievement	23	15	18	23	2	3	3	5	2	2	1	5	6	18
Affiliation	33	22	29	37	2	3	1	2	1	1	0	0	7	21
Anticipation	18	16	5	6	1	2	15	27	1	1	0	0	1	3
Celebration	10	8	8	10	1	2	1	2	0	0	0	0	2	6
Disorganisation about meeting	5	4	1	1	0	0	0	0	4	4	0	0	1	3
Fight	27	18	0	0	1	2	1	2	27	26	1	5	0	0
Goal Striving Impeded	15	12	0	0	3	5	3	05	11	10	1	5	0	0
Health of friend/relative	14	9	2	3	8	13	5	9	2	2	0	0	2	6
Helping others	14	10	4	5	4	7	2	4	7	7	1	5	1	3
Illness	9	9	3	4	4	7	4	7	0	0	1	5	0	0
Lack of time	10	7	1	1	3	5	2	4	5	5	0	0	1	3
Loneliness or isolation	6	4	1	1	4	7	2	4	0	0	0	0	0	0
Losing friendship/closeness	9	6	1	1	7	12	0	0	2	2	0	0	2	6
Morally wrong/taboo	16	11	0	0	3	5	0	0	6	6	13	59	2	6
Parting (non-dyadic)	5	4	0	0	4	7	0	0	2	2	0	0	0	0
Put down or rudeness	36	7	0	0	4	7	2	4	32	30	2	9	1	3
Role change	15	8	3	4	3	5	5	9	2	0	0	0	3	9
Unfairness-Injustice	29	21	0	0	3	5	1	2	24	23	1	5	1	3

Note. Proportions are the proportion of cases within each emotion coded at each general theme.

In some cases the specific theme labels could simply be re-phrased into sentence form to use as stems. In other cases, it was necessary to go back to the text of the interview to create the two sentence item stem representing a specific theme. An example of how this was done is shown below:

This text represents the specific theme "fight with girlfriend" (under the broad general theme Fight).

> *... I'm walking with my girlfriend travelling here* [to the appointment with the researcher]*, and she led me on this shortcut. It ended up we were on the eighth floor so we had to walk all the way down and I was blaming her - why the hell did you want to take that way when you could have just taken the other way to walk and just walked up a couple of stairs, that's more logical and then I gave her the shits as usual. Great sense of direction… Um, she wasn't too happy, I wasn't too happy either, so I just left her alone and told her I'd see her later, I'll probably give you a call later… I was like, I was all sweating, I was thinking I was going to be late and stuff and yeah, so I was quite anxious and pee'd off as a result.*
>
> This was translated to the item stem:
>
> *Hamish's partner leads him on a short-cut that takes 10 minutes longer than the normal route, making him late for a meeting. Hamish snaps at her angrily before entering the meeting.*

A similar procedure was followed to produce other item stems. For each of the 10 categories 12 item stems were created (6 representing personal life and 6 representing work life).[9] Equal numbers of items representing personal-life and work-life domains for each general theme were created, even when this equal distribution is not the case for the specific themes. Therefore, some item stems were created to correspond to a different domain. An example is given below. The first item (representing a fight in personal life) came from a reported situation, and the second was created to be analogous to this, but in the domain of work.

> *Item 1: Fight (personal life): Maryam accuses her sister Farah of making catty remarks about her. Farah responds that these remarks are true, and becomes very angry.*

---

[9] Four extra items were created for "Fight", two extra items for "Goal striving impeded" and double the amount of items (i.e., 24 instead of 12) were created for Rudeness, as so many of the reports concerned these categories

Item 2: Fight (work life): Ivy accuses one of her co-workers, Julie, of making negative remarks about her performance. Julie responds that these remarks are true, and becomes very angry.

Thus, 138 brief descriptions of situations of Anger, Sadness, Disgust, and Fear were created. These are used as item stems in the next study, where participants generate response options in the following study by replying to these scenarios.

2. Study 2: Response Option Generation

2.1 Introduction to Study 2

The primary purpose of this study was to select appropriate situations from the array developed in the previous study, collecting responses to these situations that could then be condensed to form response options. A secondary purpose was to examine relationship between personality and suggested responses, comparing participants' knowledge of the best response to their actual (self-reported) response. As in Study 1, the relevancy of situations and samples, the phrasing of response generation requests, and the summarisation of responses into response options form the basis for this study's design.

2.1.1 Selection of Situations for Inclusion as Test Items

In generating response options, the primary concerns are that they constitute a wide range of possible behaviours; that some behaviours are clearly better than others; and that the behaviours are content valid (i.e., none of the response options look obviously ridiculous or comical, as this low face validity may detrimentally affect the motivation of test-takers). The following three criteria were used to select situations that could be effectively used as item stems. Firstly, the situation must be understood by all (or almost all) participants to represent the same emotion. Secondly, all (or almost all) participants must interpret the problem presented by the situation in the same way. Thirdly, there must be adequate response variability to ensure that at least four options could be created.

2.1.2 Selection of a Relevant Sample

Ideally, the sample used would be diverse in life experience, ensuring a relatively wide range of responses be generated. However, given the practical concerns about conserving community volunteers (who are expensive and time consuming to get) it was decided that using an undergraduate student sample would not seriously compromise the validity of response generation. The inclusion of mature age students and a mix of genders were monitored to ensure that there was some variability in life experience. As in Study 1, response-option length was limited to avoid reading comprehension or verbal skill contributing too much to test-score variance. Response-option length was limited to one line of text (approximately 15 words) with the aim to have most response options as short as possible.

2.1.3 Phrasing of Response Generation Requests

Ployhart and Ehrhart (2003) found that instructions to respond to SJT items in terms of "what is the most effective response" (*best* responses) related to grade point average more strongly than "what would you do in that situation" (*actual* responses). This study did not test the effect of these different instructions on the tests' relationship to personality dimensions, although McDaniel and Nguyen (2001) suggest that actual responses will relate much more strongly to personality than best responses. Asking participants for both best and actual responses allows an empirical test of this suggestion, as well ensuring a variety of responses, ranging from very good to possibly ineffective. It is hypothesised that actual responses will relate more strongly to personality dimensions than best responses (with personality assessed by OCEANIC scores, Roberts, 2000).

2.2 Method

2.2.1 Participants

99 first year undergraduate students (56 female) at the University of Sydney participated in this study to obtain course credit. Ages ranged from 17 to 49 with a median age of 19 (30 participants were mature age students, aged 21 and over). 52 participants identified their cultural group as Australian or Anglo-Celtic, 10 as Chinese, 4 as Vietnamese, 3 as Greek and 3 as Indian (no other ethnicities were identified by more than two participants).

2.2.2 Procedure

Participants read a brief description of the experiment and emailed the experimenter agreeing to participate, after which they were emailed a form containing demographic questions, a personality measure of the five-factor model, and one of three possible forms describing several hypothetical emotional situations. For each hypothetical situation, participants were asked to describe the *best* possible response, as well as what they would do in that situation. There were three possible forms sent out to participants (Form A, B and C), described in detail below.

2.2.3 Materials

2.2.3.1 OCEANIC

The OCEANIC (Openness Conscientiousness, Extraversion, Agreeableness, Neuroticism Index Condensed; Roberts, 2000) measures five personality dimensions based on the five factor model (Costa & McCrae, 1991). There are 60 items altogether, 12 for each dimension, with Cronbach Alpha reliability indices ranging from .80 to .90 (Roberts, 2000). Participants rate how well they think each statement described them on a 6-point scale from Never to Always. *E.g., I am a kind person.*

2.2.3.2 Option Generation Form

There were three alternative versions of the option generation form (Form A, Form B and Form C) and each participant answered one of these only. Three versions were necessary so that all 138 situations could have responses generated without fatiguing the participants. Each form consisted of several situations, where participants were asked to describe both the *best* action to take in that situation, and what they would really do for each situation. Situations were organised by general theme, with at least 12 situations for each general theme (6 representing work content, and 6 representing personal life content). Table 2.6 shows the structure of each form. An example question is shown in Figure 2.2 (Section 2.3.2).

Table 2.6

Structure of Response Generation Forms Showing Number of Participants and Themes, Emotions, Domains Covered in Forms A, B, and C

Form	*N*	*% Female*	*Themes Covered*	*Emotions Covered*
A	34	16 (47%)	Fight	Anger
			Anticipation	Fear
			Put down/Rudeness	Anger
B	34	18 (53%)	Unfairness	Anger
			Role Change	Fear
			Put down/Rudeness	Anger
			Lack of others	Sadness
C	31	22 (71%)	Morally wrong or taboo	Disgust
			Goal striving impeded	Anger
			Communication	Anger
			Health concerns	Fear/Sadness

2.2.4 Coding

As in Study 1, coding for this study was undertaken using N6 software, where nodes are created, and cases may then be coded at each node. In this study, each question was analysed separately, so a "case" consisted of someone's response to the situation. 51 of the 136 situations were selected for coding as they met the criteria of being widely interpreted in the same way, to represent the same emotion and contained enough response variability to create at least four options (after coding, it was apparent that 9 of these did not meet criteria either, to make a total of 42 questions for the STEM). At this point, it was decided not to include any disgust questions, as none of these met the criteria for coding. For each situation, nodes were created to represent specific types of responses. Both the "most effective" and the "what would you do" responses were then coded as representative of more than one of these nodes.

2.3 Results and Discussion

2.3.1 Reliabilities and Descriptive Statistics for OCEANIC

Reliabilities and descriptive statistics for the OCEANIC are shown in Table 2.7, along with comparison parameters from the OCEANIC manual (Roberts, 2000). All dimensions are reliable, with alpha coefficients in the .80s, consistent with the manual. Estimates of the difference between the current sample and comparison sample are all less than .20 which Cohen (1988) defines as small.

Table 2.7

Descriptive Statistics and Reliabilities for the OCEANIC in the Current Study and a Comparison Sample from the OCEANIC Manual

	Current study			*Comparison Group (Roberts, 2000)*			*d-score*
	Mean	*SD*	*Alpha*	*Mean*	*SD*	*Alpha*	
Openness to Experience	42.39	9.53	.85	44.05	8.01	.80	-0.19
Conscientiousness	48.14	9.40	.88	47.47	8.56	.87	0.08
Extraversion	46.96	8.82	.88	47.61	9.35	.90	-0.07
Agreeableness	57.09	6.38	.84	56.56	7.01	.88	0.08
Neuroticism	35.13	8.59	.88	35.48	8.51	.89	-0.04

2.3.2 Illustrative Example of Coding, Analysis, and Option Creation for Each Item

For each situation analysed, data collection and analysis consisted of: (1) response elicitation, (2) node creation and coding, (3) personality effects on best and actual responses, and (4) creation of the final STEM item. A prototypical item analysis is shown in Figure 2.2 to illustrate this process. In this example, five "response type" nodes were sufficient to account for the 68 cases (the 34 best and 34 actual responses from the 34 participants). Frequencies were similar for "best" responses and "actual" responses, as was the case for most questions. Other questions had between 5 and 15 nodes accounting for all responses. In the example, only the first four nodes were re-phrased to use as

response options in creating the question, as it was thought the fifth ("Go back overseas") might be viewed as too extreme and hence comical by some participants.

Figure 2.2

Example of a Situation, Analysis of Responses, and Conversion of Nodes to Response Options

A. Question given to participants in Study 1

Clayton has been overseas for a long time and returns to visit his family. So much has changed that Clayton feels left out. How could Clayton deal with these feelings?

1. What is the most effective response for Clayton? _____
2. What would you do if you were Clayton? _____

B. Types of Responses, Their Frequencies, and Their Relationship to the Big Five Personality Traits

	Node Frequencies		*Differences in Personality (Cohen's d)*									
			Best					*Actual*				
	Best	*Actual*	*O*	*C*	*E*	*A*	*N*	*O*	*C*	*E*	*A*	*N*
Give time and involvement	21	18	-0.29	-0.11	0.73	-0.42	0.01	-0.88*	0.17	0.48	-0.28	-0.10
Reflect that things change	4	4	-0.04	0.73	-0.16	-0.09	0.50	-0.11	-0.27	0.09	-0.09	-0.49
Tell family how he feels	4	4	0.62	0.08	-0.72	0.08	-0.43	1.07	0.94	-0.20	0.29	-0.80
Believe that it will sort itself out	4	5	-0.24	-0.42	-0.30	0.60	-0.47	0.27	-0.04	-1.06**	0.36	0.31
Go back Overseas	1	3	-	-	-	-	-	0.30	-0.52	-0.64	0.51	1.39**

Note. T-tests were run to check significance of differences in personality scores for participants suggesting or not suggesting a particular response

*$p < .05$, **$p < .01$

C. Final Multiple-Choice Item Created from the Responses Generated

Clayton has been overseas for a long time and returns to visit his family. So much has changed that Clayton feels left out. *What action would be the most effective for Clayton?*

a. Nothing, it will sort itself out soon enough
b. Tell his family he feels left out.
c. Spend time listening and getting involved again.
d. Reflect that relationships can change over time.

Figure 2.2 also illustrates decisions made regarding explicitly stating the goals of the situation's protagonist. Essentially, goals *were* explicitly stated in this current study, but were left up to inference in the final STEM items. For example, while Clayton "deal[s] with his feelings", other protagonists make peace with their partners, manage their grief, or deal with their nerves. This

specificity was chosen to help participants interpret situations in the same way. However, in the STEM items, questions were phrased in the same manner for all items: "What is the most effective response for [protagonist's name]?" Although this meant the protagonist's goals were inferred (and this inference therefore an additional source of between-subject variance), there were two reasons this format was chosen. Firstly, standardising the question over all items is good psychometric practice for items assessing the same construct. Secondly, the additional variance *is* likely to be construct-related variance, insomuch as effective management of emotional situations requires inferring or creating goals rather than just selecting the best way to achieve them.

Similar analyses were undertaken for all 51 items selected for coding, and these are shown in Appendix 2.4. However, response options for the STEM were only created for the 42 situations that met appropriate criteria. Of these, 18 were anger-related (12 work-related, 6 from personal life), 12 were fear-related (6 work-related, 6 from personal life), and 12 were sadness-related (4 work-related, 8 from personal life). In each emotion-by-situation category, half the situations had male protagonists and half had female (it was necessary to change the gender in four situations in order to achieve this balance).

2.3.3 Relationship of "Best" and "Actual" Responses to Personality Traits

In the example in Figure 2.2, the relationship to personality traits differed for best and actual responses. Mean personality scores for each of the five personality dimensions were calculated separately for groups suggesting and not suggesting each node, where Cohen's effect size (*d*) estimated the difference between these groups. Extraversion's effect size was .73 for the "give time and involvement" node for best response, indicating that participants suggesting this response were more extraverted than those not suggesting it. In the example, the only significant differences between personality means occurred for the "actual" responses. People giving time and involvement were significantly less Open to experience, people believing situation would sort itself out were significantly less Extraverted, and people who would go back overseas were significantly more

Neurotic. This is as expected, since "actual" responses were predicted to be more strongly related to personality than "best" responses.

Over all situations, best and actual responses differed by 10%, and this value was the same, to one percentage point, for male and female participants (i.e., ideal and actual responses different only 10% of the time). The small size of this difference may be partly due to the methodology used, where participants were asked for best and actual responses consecutively, and thus thinking of the best response first may have influenced what they proposed as their actual response. It is probable that greater differences in response would be found if these were not contiguously placed.

However, even with small differences between best and actual responses, there were significant differences in the strength of relationship to personality traits. The mean effect of personality on response choice over all five personality traits (in terms of *d*-scores) was 0.39 (*SD* = 0.33) for best responses and 0.44 (*SD* = 0.37) for actual responses, and this difference was significant ($t_{2277} = 3.059$, $p = .002$). The mean effect of personality was also significantly stronger for Extraversion (mean $d = 0.46$, $SD = 0.39$) than Neuroticism (mean $d = 0.40$, $SD = 0.35$) under multiple comparisons over all five personality dimensions with Tukey's honestly significant difference set at $p < .05$. No other differences between personality traits were significant.

Since personality effects were stronger for "what would you do" phrasing of the question, it was decided not to use this phrasing in the STEM items. For each of the 42 questions, participants were asked "what action would be most effective for [name of protagonist]".[10] The studies below describe the development of two scoring keys in two different ways: (1) according to expert judgment (see Study 3), and (2) by summarising the literature on coping effectiveness in different types of situations (see Study 4).

[10] In addition to the 42 questions developed this way, 2 additional questions representing sadness were introduced, as there were concerns about having less than 5 questions to represent sadness in a work domain. Thus the final version of the STEM contains 44 questions.

3. Study 3: Expert Judgment of the Correctness of Situations

3.1 Introduction to Study 3

The third study represents the last step in the development of an SJT, involving the expert judgment of the response options to determine scoring weights. Since the STEM will be administered in both multiple-choice and ratings-based judgment formats, expert judgment was collected in both multiple-choice and rating-scale formats.[11] This study's purpose is to derive an expert-based scoring key, based on the convergence of the sample's responses around an ideal point. In line with concerns that there are multiple possible domains of expertise, experts were recruited from a number of different domains: expertise in declarative clinical knowledge (senior clinical PhD students); expertise in declarative knowledge of emotions and emotional intelligence (members of an Australian EI research group); expertise in practical problem solving in typical day-to-day situations (life coaches); and expertise in emotional problems in a clinical field (therapists, counsellors, and clinical psychologists). Separate samples of experts were used to generate scoring rubrics for multiple-choice and ratings-based judgments formats of the test.

3.1.1 Collection of Expert Weights

Expert weights may be applied to the *multiple-choice* STEM in two ways: (1) proportionally, or (2) with mean expert ratings. In proportion scoring, experts complete the test in multiple-choice format and the proportion selecting each option is used as a scoring weight (as in the MEIS and MSCEIT). In mean rating scoring, experts rate each option, and the mean of these ratings is used as a scoring weight (e.g., if the average expert rating of option A in question 1 was 4.8 out of 6, a test-taker endorsing option A would receive a score of 4.8). Expert weights may be applied to the *ratings-based* STEM in two ways: (1) by proportion of experts endorsing each scale-point (as in the MEIS

[11] Although it would be possible to score the multiple-choice test based on the expert ratings of options (awarding the expert mean to the option selected, or awarding 1 point for choosing the option rated the most highly by the experts), it was felt that it would be useful to collect expert data in multiple-choice format also, as this is the way the current tests of performance-based EI (the MEIS and MSCEIT) are scored.

and MSCEIT); or (2) by the distance between participants' ratings and an expert mean. Distance scoring is commonly used in social intelligence and tacit knowledge tests (e.g., Hedlund et al., 2003; Legree, 1995; note that there are numerous possible ways of computing distance scores). Thus there are *three* sets of expert weights that this study will produce: (1) expert proportions for each option; (2) mean expert ratings for each option; and (3) expert proportions for each scale point of each option.

3.1.2 Validity of Expert Judgments

In addition to collecting expert judgments, this study also examines the validity of the expert judgments. Four issues are addressed: (1) that agreement should be higher for experts than non-experts; (2) that experts with low agreement indices be excluded from scoring rubrics; (3) that experts score more highly than non-experts; and (4) that gender differences are not due to the gender of the expert sample. The non-experts used to examine issues (1), (3), and (4) were the sample used in Chapter 3 of this thesis (see Method [Section 2.1] of Chapter 3 for more detail about this sample).

(1) Higher agreement among experts than non-experts. Given that judgments from non-expert groups have more error (i.e., random variation) than experts (Legree, 1995; Legree et al., 2004), the current expert group ought to agree more strongly among themselves as to the correct answers than do non-experts. Agreement indices can be calculated for the expert group and non-expert group to determine whether the expert group has a higher level of agreement (and hence can validly be defined as experts).

(2) Exclusion of experts with low levels of agreement. If agreement among experts is indeed higher than agreement among non-experts, then agreement with other experts can be taken as a sign of expertise. Each expert can be screened for their agreement with the others, and excluded if agreement indices are too low.

(3) Expert scores are higher than non-expert scores. The expert mean on the STEM should be significantly higher than the non-expert mean. If experts do not score significantly more highly than non-experts, then their status as experts in emotion management is not valid.

(4) Gender differences among non-experts should not be due to the gender of the expert sample. Roberts et al. (2001) suggest that characteristics of expert scorers may lead to group differences in test scores, as evidenced by the higher scores of men on the MSCEIT, where experts were male. To test this effect, gender differences in STEM scores (among non-experts) can be compared over two cases: (1) when scored by expert weights determined from male experts only; versus (2) when scored by expert weights determined from female experts only. If males do relatively better under male experts, and females under female experts, this is evidence that expert characteristics are responsible for group differences. The adequacy of the expert group can thus be assessed reasonably thoroughly, in several ways, with the judgments of doubtful experts removed before expert-scoring keys are created.

3.2 Method

3.2.1 Participants

Expert samples. 18 experts participated in the study. 13 experts (9 female) completed the multiple-choice version of the STEM (ages ranged from 25 to 64; $M = 43.1$, $SD = 13.9$). 6 experts (all female) completed the ratings-based judgment version (ages ranged from 25 to 51; $M = 39.4$, $SD = 9.4$). One of the members of the EI research group completed both versions. For the multiple-choice version, three experts were post-graduate clinical psychology students in the final or penultimate year of their course (which involved clinical placements and academic training); three were members of an Australian EI research group; two were counsellors; one was a psychologist; and four were life-coaches. For the ratings-based judgment format, there were three coaches, a school psychologist, a psychotherapist, and a member of the EI research group.

Non-expert sample. A non-expert sample was taken from the study outlined in Chapter 3. 112 participants completed the STEM in multiple-choice format and 91 in ratings-based format. Further information on this sample is provided in the Section 2.1 of Chapter 3.

3.2.2 Materials

Expert volunteers were given a copy of the STEM to complete in their own time and return to the experimenter by mail or email (depending on their preference).

3.3 Results and Discussion

3.3.1 Multiple-Choice Format: Expert Agreement

Kappas and raw agreement indices over all 44 questions of the STEM were calculated in order to estimate of the level of agreement between experts (these indices are shown in Appendix 2.5). Expert 11 had a much lower level of agreement than the others, with some Kappa estimates as low as .09. This individual's judgments were thus not included in expert scoring weights. Raw levels of agreement ranged from .32 (or .45 if expert 11 is excluded) to .73, averaging .57 (*SD* = .09), where the chance rate would be .25. Kappas ranged from .09 (.26 if expert 11 is excluded) to .63, averaging .41 (*SD* = .12). In a randomly drawn sample of 13 non-experts from the validation study in Chapter 3, mean raw agreement was .54 (*SD* = .07) and the mean Kappa was .38 (*SD* = .09). Both of these differences were significant ($t_{154} = 2.14$, $p = .034$ for raw agreement; $t_{154} = 2.115$, $p = .036$ for Kappa).[12] When expert 11 was excluded (and the non-expert with the lowest levels of agreement with other novices was excluded), these differences became greater. This significantly greater convergence among the experts (compared to the non-experts) is an indication of their greater expertise. Thus there is evidence that expert selection was valid.

3.3.2 Multiple-Choice Format: Expert Superiority to Non-experts

When determining experts' scores on the STEM it is important that expert scoring weights are not determined from the same sample that these weights are then used to judge (or expert scores will be over-estimated, see Cureton, 1950). Therefore, expert performance on the multiple-choice STEM

[12] Six groups of novices were selected at random and the differences between within-novice and within-expert agreement was significant for all six groups.

was scored according to the mean rating of the experts completing the ratings-based form of the STEM (and the expert who completed both forms was excluded from the multiple-choice sample).[13]

Expert scores on the multiple-choice STEM (scored by mean ratings of the second group of experts) were compared to non-experts scores of 112 participants who completed the multiple-choice STEM in Chapter 3 (also scored by mean ratings of the second group of experts). The expert mean (M = 4.61; SD = 0.14) was significantly higher than the non-expert mean (M = 4.47; SD = 0.19; t_{122} = 2.51, p = .013), indicating that experts score more highly than non-experts on this test. However, the effect size was small (d = .21), in Cohen's (1988) terminology.

3.3.2 Multiple-Choice Format: Does the Gender of the Experts Cause Gender Differences in Test-Takers' Scores?

In this analysis, the STEM was scored according to the proportion weights of the experts who completed the multiple-choice STEM. More specifically, two possible sets of expert weights were derived: one from the four male experts only (male weights) and one from the nine female experts only (female weights). The performance of the female non-experts (n = 77) and male non-experts (n = 35) from Chapter 3 were compared under both male weights and female weights. Females scored higher than males under female weights, but not significantly so (female mean = 23.58, SD = 2.59; male mean = 22.50, SD = 3.90; t_{110} = 1.58, p = .117). However, females' scores were also non-significantly higher under the male expert rubric (female mean = 21.30, SD = 2.50; male mean = 20.21, SD = 3.26; t_{110} = 1.93, p = .057), and this difference was in fact slightly larger. Thus, there is no evidence that the characteristics of the experts lead to group differences. This is not to say that group differences do not exist, but merely that they do not appear to be due to the composition of the expert group.

[13] One ratings-based expert did not contribute to these expert mean rating. She was excluded due to low agreement with the other experts; see the section "Ratings format: expert agreement".

3.3.3 Ratings Format: Expert Agreement

For the ratings-based judgment format, agreement was calculated in two ways: (1) the Pearson correlation, over all questions, of each expert's ratings with each other expert's ratings; and (2) the mean distance of each expert from the other five experts, averaged over all questions. These are also shown in Appendix 2.5. Expert 1 had the lowest mean correlation and their distance scores were significantly different to experts 1 and 2, 5, and 6 using Tukey's honestly significant difference to estimate multiple differences at $p < .05$. For this reason, the judgments of expert 1 were not included in developing the ratings-based scoring key.

Pearson correlations and mean distances were also calculated for six participants randomly selected from the study described in Chapter 3. Whereas experts had a mean correlation of .61 ($SD = .05$), non-experts had a mean correlation of .37 ($SD = .21$). This difference was significant ($t_{16} = 4.34$, $p <. 001$). Similarly, the mean expert distance was 1.06 ($SD = 0.09$) which differed significantly from the mean non-expert distance of 1.28 ($SD = 0.14$; $t_{10} = 3.22$, $p < .01$). That is, experts' responses on the ratings-based judgment format converge more closely than non-experts, which bodes well for the validity of selecting experts. As there were only five valid experts for the ratings-based judgment format of the STEM, and all were female, it was not feasible to analyse these for expert/non-expert differences or gender differences.

Appendix 2.6 shows the three sets of expert weights created from this study: (1) proportions of experts selecting each option; (2) proportions of experts selecting each scale point for each option; and (3) the mean expert rating of each option. These three sets of weights are used to score the STEM in Chapters 3 and 4.

4. Study 4: Development of a Standards-Based Scoring Key for the STEM

In addition to a scoring key based on expert judgment, a set of guidelines for scoring based on published theory and evidence was also constructed. The summary of empirical evidence is given in Appendix 2.7 and the final scoring key in Table 2.8. This set of guidelines will be applied to each option of the STEM, where more effective strategies are weighted more highly than less effective

strategies for each item. Table 2.9 outlines how these guidelines were applied in one particular sadness-related item.

Table 2.8 lists several coping strategies in order of effectiveness for five different types of situations: (1) Chronic illness; (2) One-off or acute illness; (3) Work-related situations; (4) Loneliness; (5) Other interpersonal situations. The order of effectiveness is based on a summary of empirical evidence given in Appendix 2.7. To apply this scoring key, each situation must be categorised as one of these five types of situations, and then each response option categorised as one of the coping strategies that shows differential effectiveness in these situations.

Table 2.8

Guidelines for Determining the Effectiveness of Coping Strategies in Different Types of Situations

Illness (Chronic)
Effective Strategies: problem-focused (information seeking); positive re-appraisal (appraisals of control); distancing from pain; versatile coping; seeking social support *Ineffective Strategies*: wishful thinking and/or self-blame (i.e., forms of emotion-focused coping); avoidance
Illness (one-off or acute)
Most effective: combination of emotion- and problem- focused coping *Effective*: active behavioural coping; relaxation; social support; passive coping (i.e., acceptance or not using strategies); positive reappraisal *Mixed or conflicting evidence for effectiveness*: avoidance or ignoring of pain; attention to physical sensations *Ineffective:* emotion-focused; attention/focusing on feelings
Work-related situations
Most effective: perceived control over negative moods; perceived control over situation (i.e., appraisals of challenge rather than threat); problem-focused coping *in conjunction with distancing or seeking social support*; problem-focused coping *when the situation is clear* or *when the person has sufficient authority to act*. *Effective*: Distancing; Problem-focused; Seeking Social Support *Ineffective:* acceptance combined with venting; avoidance; emotion-focused coping; appraisals of threat; appraisals of irrelevance (if the task is to solve the problem, rather than ameliorate negative emotions).
Loneliness
Effective: approach strategies; acceptance and reflection; social interaction; increased activity *Ineffective*: avoidance strategies; distancing; attempts to ignore loneliness; attempts to deny loneliness.
Everyday and Interpersonal Situations
Effective: problem-focused coping; positive re-appraisal (except for medium and low threat, where this increases anxiety); seeking social support *Ineffective:* confrontive/hostile coping; avoidance; planning with action (i.e., rumination); self-blame; wishful thinking; accepting the blame.

An example of how these guidelines were applied in shown in Table 2.9 for a sadness item. Since this item represents the "Loneliness" type of situation, the option representing avoidance is clearly the worst. The fourth option ("tell his friends he is disappointed in them …") does not obviously represent any of the coping strategies for Loneliness, but does represent two of the ineffective strategies for coping with an interpersonal situation (rumination confrontation). These two responses are therefore the worst, and selecting either of these would score no points. Both A and B represent coping strategies that would be effective in a loneliness situation, and these therefore score 1 point. An outline of the decisions made in this way is given in Appendix 2.8 for each of the questions on the STEM.

Table 2.9

Example of an Application of the Coping-Based Scoring Guidelines to a STEM Item

	Option Represents	*Score*
Andre moves away from the city his friends and family are in. He finds his friends make less effort to keep in contact than he thought they would.		
[A] He should make the effort to contact them, but also try to meet people in his new city.	Approach, social interaction, increased activity, plus it's versatile.	1
[B] Try to adjust to life in the new city by joining clubs and activities there.	Increased activity and social interaction.	1
[C] Let go of his old friends, who have shown themselves to be unreliable.	Distancing or avoidance.	0
[D] Tell his friends he is disappointed in them for not contacting him.	Somewhat related to rumination, plus it is confrontive.	0

Part III of Chapter 2

Chapter Summary

1. Overview of Test Development of the STEU and STEM

This chapter outlined the process by which the Situational Test of Emotional Understanding (STEU) and the Situational Test of Emotion Management (STEM) were created as measures of Understanding and Management respectively. The STEU is a multiple-choice test assessing understanding of emotions in work contexts, personal-life contexts and no context over 14 emotions. The STEU is scored according to Roseman's (2001) empirically-derived theory. STEU items are used in Chapters 3, 4, 5 and 6. Chapter 3 divides the STEU into 3 components according to the context of the items, whereas Chapter 4 divides the STEU into 14 item parcels according to the 14 emotions covered by the test.

The STEM may be given in either multiple-choice or ratings-based format and each of these formats may be scored either by a coping-based scoring key or by two sets of expert weights (the mean expert rating of each response, or the proportion of experts selecting each response option as the best). The STEM measures emotion management in work-life and personal-life contexts for the emotions sadness, fear, and anger. Situations represent one of eight general content areas. Fear items are drawn from three content areas: (1) Anticipation, (2) Role Change, and (3) Health Concerns. Anger items are also drawn from three content areas: (4) Goal Striving Impeded, (5) Unfairness, and (6) Put-down/Rudeness. Sadness items are drawn from two content areas: (7) Isolation and (8) Parting. The division of the STEM into three emotion components is undertaken in Chapter 3, and the division of the STEM into item parcels by these general content areas is undertaken in Chapter 4.

Results from Study 1 suggested that joy, disgust, and surprise should not be included in the STEM's content for the following reasons. Joy was person as well as situation related (i.e., whether joy occurred was due to individual differences as well as situational characteristics) and was felt at significantly weaker intensities than other emotions. Surprise tended not to occur alone, but in

combination with other emotions (particularly sadness and joy). There were too few disgust-related situations to develop clear items. This meant that the STEM covered the emotions anger, fear, and sadness only.

Study 2 found participants responded differently to the same situations when given different instructions: "what is the best thing to do" (best), and "what would you do in that situation" (actual). Actual responses related significantly more strongly to personality than best responses. Of the five personality dimensions, Extraversion affected participants' responses the most and Neuroticism the least (i.e., suggested responses to emotional situations were significantly more influenced by Extraversion than Neuroticism).

Study 3 confirmed that experts used to create an expert-scoring-key for the STEM really did have higher levels of expertise than non-expert first year psychology students. The experts agreed more strongly amongst themselves than the student sample, and scored significantly more highly on the STEM than the student sample. In addition, there was no evidence that the gender composition of the expert sample resulted in gender differences in the student sample's expert-scored STEM results.

Study 4 provided an alternative to expert scoring of the STEM with a possible standards-based scoring key. Evidence of different levels of effectiveness for different types of responses for each of five types of situations (Chronic Illness, Acute Illness, Work-Related Situations, Everyday/Interpersonal Situations, and Loneliness) may be used to score items according to the similarity of these items to the five types of situations. This method of scoring is used in Chapter 3.

2. Limitations of Test Development of the STEU and STEM

Limitations of Content Coverage

Both the STEU and the STEM's content covered limited number of emotions. The STEU covered fourteen emotions, and the STEM three (anger, fear, and sadness). Care was taken *not* to present blended emotions (i.e., a combination of two or more emotions). In addition, although items represented different levels of emotional intensity, intensity level was not clearly built into the structure of the test to allow examination of the effect of intensity (e.g., as would be the case if half

the items for each emotion represented high intensity and half the items low intensity). Emotion type, blends of emotion, and intensity of emotions might all be relevant factors to control or examine in the study of emotional Understanding and Management. These questions might be addressed in future research or future test developments of socio-emotional skills or abilities, but are beyond the scope of this thesis.

Limitations of Samples

Although considerations of what constitutes a relevant sample for emotion management were adhered to, the samples for each of the three stages of STEM test construction were not perfect. Both situations and responses were generated from a relatively narrow range of people. Situations were generated by a sample consisting of a large proportion of psychology students (62% of the participants were students) and only people who lived or worked in Sydney. The specific details of the situations this sample experienced might well differ from rural, non-tertiary educated people. Responses to the situations were generated by a sample consisting entirely of psychology students, compounding this issue. Students might thus find the content of the STEM more familiar and relevant than others. This is a primary reason why Chapter 4 examines the properties of the STEM and STEU in a non-student sample. Just as the content generation samples were skewed towards psychology students, the expert samples were skewed towards females. Three quarters of the experts used for the multiple-choice STEM were female, and all of the experts used for the ratings-based scoring key were female. Whether this female majority among expert samples relates to a spurious female superiority on these tests is examined empirically in further chapters.3.

3. Use of the STEU and STEM in Subsequent Chapters

In the chapters that follow, these tests will be used to address some important questions regarding individual differences in EI. These questions include the effect of different response formats and scoring types on tests' reliability and validity, the usefulness of EI in predicting emotion-related criteria, the location of EI within Gf/Gc theory, the relationship of EI to state and trait distress and EI's incremental prediction of achievement.

Chapter 3

Examining Methodological Issues in EI Assessment and the Validation of the Situational Test of Emotional Understanding (STEU) and Situational Test of Emotion Management (STEM)

1. Introduction to Chapter 3

This chapter has two primary aims: (1) a comparison of response-formats, and (2) test validation of the STEU (Situational Test of Emotional Understanding) and STEM (Situational Test of Emotion Management). Two response formats are directly compared by administering the STEM to separate groups in different formats: (1) ratings-based judgments; and (2) multiple-choice. After examining whether there are individual differences in ratings-scale use (which may confound construct measurement) the effect of each response format on the reliability, validity, and utility of the STEM (Situational Test of Emotion Management) will be examined. In addition, evidence of convergent, discriminant, and concurrent validity of the STEU and STEM will be evaluated by comparing these scores to measures of EI, intelligence, personality, alexithymia, life satisfaction, psychological distress, and academic achievement. These issues are discussed in further detail in the paragraphs below.

1.1 Test Construction Issues: Response Format and Scoring Type

1.1.1 Individual Differences in the Rating-Scale Use: Examination and Suggested Solution

All but two tests in the MSCEIT (and the MEIS) battery ask test-takers to *rate* stimuli (e.g., rate the presence of an emotion from 1 to 5, rate the effectiveness of a management strategy from 1 to 5; Mayer et al., 1997, 2002b). Allowing test-takers to rate each option elicits more information than getting them to select one best answer, which is presumably why the MEIS and MSCEIT's test authors chose to make most tests ratings-based rather than multiple-choice. However, a downside to this ratings-based format is the possibility that individual differences in the use of a rating scale may

be confounded with the construct of interest (EI). Defensive people may give low ratings to all response options, whereas suggestible or acquiescent people may rate all responses highly. This might particularly be the case when scale points are not anchored with descriptors at each point of the scale, as implied by higher reliabilities in anchored versus non-anchored personality scales (e.g., Weng, 2004). Only three of the eight MSCEIT tests are anchored at each point, meaning that individual differences in rating-scale use are a very real possibility. To avoid this, the ratings-based version of the STEM gives descriptors at each point of the 6-point scale to (1 = *Not at all effective*, 2 = *Not very effective*, 3 = *Slightly effective*, 4 = *Somewhat effective*, 5 = *Fairly effective*, 6 = *Extremely effective*). Even so, this chapter examines the possibility that meaningful individual differences in the use of the rating scale persist. Individual differences in rating-scale use will be examined for two scales: (1) the Situational Test of Emotion Management (STEM), which uses anchored scale points (see Part 2 of Chapter 2, for a description of the STEM's development), and (2) the Stories test from the MEIS, which anchors only the end points on the scale (Mayer et al., 1997).

If individual differences in the use of the rating scale do exist, this may lower the validity of scoring mechanisms that assume rating-points have the same meaning across individuals. The possibility of males and females using the rating-scale of the Stories test differently was examined by Mayer et al. (2000b) who found no significant difference between genders. The current study expands on Mayer's examination, exploring differences between experts and novices, differences based on cultural identification as well as gender, and also comparing individual differences in rating scale use to personality variables.

If each individual is consistent in the way they apply the rating-scale over all items, there are methods that partial out individual differences in rating-scale use. One possibility proposed by Legree (1995) is to first standardise the ratings over each individual, and then use these standardised values when calculating the distance from the ideal or "correct" value. Other methods include correlational scoring which compare test-taker's profiles of option ratings in a given vignette to ideal profiles (e.g., Pearson or Spearman methods; see Legree et al., 2004). A potential problem with correlational

scoring is that it cannot be used for participants who rate all options equally on a given vignette (i.e., a correlation cannot be calculated for a constant). Legree's standardised scoring does not suffer from this problem, and thus will be the method of profile scoring used should there be individual differences in rating-scale use on the STEM or Stories.

1.1.2 Comparison of Response Formats: Multiple-Choice Versus Ratings-Based

1.1.2.1 Response Formats and the Relationship to Intelligence

The performance-based measures of EI currently in circulation (i.e., the MEIS and MSCEIT; Mayer et al., 1997; 2002b) use only ratings-based judgment response formats for tests in Perception, Facilitation, and Management branches, whereas the Understanding tests are either entirely or partly multiple-choice (for the MSCEIT and MEIS, respectively). Prior findings that Understanding relates more strongly to intelligence than the other branches do may be due to this response-format issue.[14] That is, Management, Perception, and Facilitation tests might also relate strongly to intelligence if they were presented in multiple-choice format. Some support for this proposition is provided by Roberts et al. (submitted), who found MSCEIT perception tests were unrelated to Gf and Gc, but that multiple-choice measures of facial expression and vocal expression recognition related *did* relate to Gf and Gc (more strongly to Gf).

The *lack* of a relationship of EI measures with intelligence may thus be an epiphenomenon or method effect of the use of a ratings-based response format. Social intelligence (SI) has much theoretical overlap with EI and has been measured by very similar items, yet has consistently failed to be differentiated from academic intelligence (Hedlund & Sternberg, 2000; Kihlstrom & Cantor, 2000; Landy, 2005, in press; Riggio, Messamer, & Throckmorton, 1991). Whereas the MSCEIT and MEIS are almost entirely composed of ratings-based scales, the two most commonly used tests of social intelligence are not (these tests are The George Washington Social Intelligence Test; Moss, Hunt,

[14] However, note that the MSCEIT's Understanding tests are still consensus-scored rather than standards-scored, unlike intelligence tests.

Omwake, & Woodward, 1955; and the O'Sullivan and Guilford Social Intelligence tests; O'Sullivan & Guilford, 1976). This provides some support for the idea that the idiosyncratic response format may be what distinguishes EI from other forms of intelligence.

Since the STEM has been constructed in *both* multiple-choice and ratings-based format, the effect of response format on EI-intelligence correlations can be empirically examined, which forms an important focus for this chapter. The relationship between Management and an intelligence marker (vocabulary) can be assessed separately for ratings-based and multiple-choice versions of the STEM. If the higher intelligence correlations for Understanding compared to Management are unrelated to response format, then the correlation between STEM scores and vocabulary should be the same whether the STEM is ratings-based or multiple-choice. However, if the higher correlation between Understanding and intelligence is due to the multiple-choice format of Understanding tests, then the STEM should correlate more highly with vocabulary when it is in multiple-choice format. This state of affairs is hypothesised to be the case: Correlations with vocabulary will be significantly higher for the STEM when it is administered in multiple-choice rather than ratings-based format.

1.1.2.2 Response Formats and the Distinction between Understanding and Management

The distinction between Understanding and Management might also conceivably be partly response-format-related rather than construct-related. To reiterate, all empirical support for the four-branch structure of EI is derived from the MEIS and MSCEIT test batteries, where Understanding tests are entirely multiple-choice (for the MSCEIT) or partly multiple-choice (for the MEIS) and all Management tests are in ratings-based format. If Understanding and Management tests were given in the same response format, they might be more strongly related.

This chapter tests the distinction between Understanding and Management with structural equation models. Nested models testing the differentiation of Understanding and Management will be compared for two cases: (1) when both Understanding and Management indicators are in the same format (multiple-choice); and (2) when Understanding and Management indicators are in different formats (Understanding as multiple-choice and Management as ratings-based). If the response format

is responsible for the differentiation of Understanding and Management, then Understanding/Management relationship will be considerably stronger when tests are in the same response format. An effect of response format is hypothesised, with Management and Understanding factors expected to correlate more strongly when items utilise the same response format than when they use different response formats.

1.2 Test Validation of the STEU and STEM

1.2.1 Convergent Validity Evidence: Relationships between EI Measures

The most important indication of convergent validity of an EI test is its relationship to other EI tests. The STEU and STEM should clearly correlate strongly if these are valid tests of EI, as they both measure Strategic EI. However, both STEU and STEM should also relate to other areas of EI, and the Stories test (from the Perception branch of the Experiential area of the four-branch model) was included to establish the STEU and STEM's convergent validity in relation to a lower branch and area. Roberts et al. (2001) found Stories showed the highest loading on a MEIS general factor and is thus arguably the best *global* measure of performance-based EI. Thus, STEU and STEM scores are hypothesised to correlate with Stories, but not as strongly as they correlate with each other.

1.2.2 Convergent and Discriminant Validity Evidence: Relationships with Intelligence

One important set of criteria for EI to attain the status of an intelligence is the correlational criteria of shared and unique variance (Matthews et al., 2002; Mayer et al., 2000b, 2001; Roberts et al., 2001). If EI tests are merely an alternative interface for the measurement of cultural and verbal knowledge, then the construct is not new but rather an alternative method for measuring Gc. Nevertheless, if EI tests were completely unrelated to existing measures of intelligence, the use of the term "intelligence" would not be justified (i.e., low correlations with intelligence would indicate poor convergent validity, Mayer et al. 2000b; 2001; Roberts et al., 2001). Thus, although EI should relate to other intelligences there ought to be some unique variation shared only by the EI tests.

In the current study, a vocabulary test (representing the primary mental ability of lexical knowledge, and the broad ability of crystallized intelligence) was used as an intelligence indicator.

Vocabulary measures have very often been used in research involving the MEIS and MSCEIT, such that EI-intelligence relationships in the current study may be directly compared with prior findings from this research corpus (see Barchard, 2003; Lam & Kirby, 2002; Lopes et al., 2003; MacCann et al., 2004; Mayer et al., 2000b; and Roberts et al., 2001, all of whom used vocabulary tests as criterion measures of intelligence, and Mayer et al., and Lopes et al. who used vocabulary as the *only* intelligence measures). Additionally, the most obvious source of shared method-related variation in the Stories, STEU, and STEM (all of which require reading of paragraphs of text) is their dependence on verbal ability. Analyses accounting for vocabulary can assess whether there is additional shared variance among the EI tests due to EI-related rather than vocabulary-related factors. Essentially, it is hypothesised that STEU and STEM scores will show significant positive correlations with vocabulary test scores, but that EI tests (STEM, STEU, and Stories) will display unique EI-related shared variance, correlating with each other after vocabulary-related variance is controlled for.

1.2.3 Convergent Validity Evidence: Relationships with Alexithymia

Alexithymia and its conceptual and empirical links to EI are described in Section 6.2.3 of Chapter 1. The Toronto Alexithymia Scale (TAS-20) used in this chapter assesses three key components of alexithymia: (1) difficulty identifying one's own emotions, distinguishing between them, and distinguishing emotions from physical sensations such as hunger (DIF); (2) difficulty describing or expressing one's own emotions to other people (DDF); and (3) an externally-oriented thinking style geared to concrete, external stimuli rather than emotional or contextual information (EOT; see e.g., Parker, 2004).

The first two of these components conceptually map onto the Experiential EI area of the four-branch hierarchical model (Mayer et al., 2001). Mayer and Salovey's (1997) detailed description of the Perception component of EI involved both the identification of emotion in physical states, feelings, and thoughts (conceptually related to identifying emotions or DIF) and the accurate expression of emotions and emotional needs (conceptually related to describing emotions or DDF; see Table 1.1 in Chapter 1 for an detailed conceptual breakdown of the four-branch model). In

contrast, the third component of alexithymia (externally-oriented thinking or EOT) seems to involve a cognitive component, resembling a style of thinking that considers the emotional details of a situation as more important data to analyse or cognise than the non-emotional details. This component of alexithymia seems more likely to map onto the strategic area of the four-branch hierarchical model of EI, which requires higher-level processing of emotion-related information (again, see Table 1.1 of this thesis). Personality correlations differ for the first two components and EOT, giving some support to the idea that EOT may be more cognitive than the other components. Bagby, Taylor, and Parker (1994b) found that DDF and DIF related more strongly to Neuroticism and Extraversion (particularly Depression and Vulnerability facets) whereas EOT related more strongly to Openness, which has been linked to crystallized intelligence (e.g., Ackerman & Heggestad, 1997; Gignac, 2005; Gignac et al., 2004).

In previous empirical studies involving the MSCEIT, both Understanding and Management have been linked with (less) difficulty identifying and describing emotion, although magnitudes were small (Barchard & Hakstian, 2004; Warwick & Nettelbeck, 2004). Neither of these studies examined externally-oriented thinking (which was excluded due to a lack of reliability in Barchard & Hakstian's study). Although the STEU and STEM are expected to relate to alexithymia generally, higher correlations with EOT than with the other two components of alexithymia might be expected, as they are measures of Strategic EI.

1.2.4 Convergent Validity Evidence: Relationships with Life satisfaction

Understanding emotional situations as well as managing them effectively through use of more and better strategies ought to result in greater satisfaction with one's life. Existing research has found small but non-trivial relationships between life satisfaction and performance-based measures of Management and Understanding in the range of .10 to .25 (Bastian et al., 2005; Ciarrochi et al., 2000; Mayer et al., 2000b). In the current chapter, this relationship will be explored for the two new measures of Strategic EI: the STEU and the STEM. In addition, a methodological extension on some

prior life satisfaction research will be implemented, by considering life satisfaction both as satisfaction with the current state of one's life, and as satisfaction with life up until the current point.

When McDonald (1999) factor analysed the Satisfaction With Life Scale (SWLS; Pavot & Diener, 1993; used in Bastian et al.'s [2005] analysis) he found two separable constructs: a retrospective perspective (i.e., satisfaction with the past events or conditions of one's life), and a current/non-retrospective perspective (i.e., considering satisfaction with the current conditions of one's life). The current investigation divides life satisfaction into retrospective and current components and before examining whether the STEU and STEM relate to life satisfaction. Those with higher emotion management may be more likely to use positive re-appraisal as an emotion regulation strategy (Lyons & Schneider, 2005) and hence may have higher retrospective life satisfaction. In addition, Salovey et al. (2000) suggest that those with high EI will have lower levels of rumination, dwelling on past pain and failure less than people with low EI. Consequently, it seems likely that Strategic EI may show stronger relationships to retrospective life satisfaction than to current life satisfaction, and that this relationship will be stronger for the STEM than the STEU (since underlying mechanisms involve the use of emotion regulation strategies). It is thus expected that the STEU and STEM will show small positive relationships with life satisfaction in line with prior research, and that this may be stronger for STEM and retrospective satisfaction.

1.2.5 Convergent Validity Evidence: Relationships with Psychological Distress

Several researchers have suggested a theoretical link between emotional intelligence and better mental health (Ciarrochi et al., 2002, 2003; Matthews & Zeidner, 2000; Salovey et al., 2002; Slaski & Cartwright, 2002, 2003). Salovey et al. (2000) suggest that high EI leads to better coping with negative emotions, which lessens psychological distress. If the STEU and STEM are to be useful measures of EI, then they should predict lower states of psychological distress. This relationship might be expected to be stronger for the STEM than the STEU, as the STEM explicitly measures knowledge of how to manage situations involving negative or distressing emotions (rather than just understand them). Psychological distress is assessed here with the state version of the General Health

Questionnaire (20 item version; GHQ-20), which assesses the symptoms of psychological distress over the previous two week period.

1.2.6 Discriminant Validity: Relationships with Personality

To ensure independence from personality, the STEU and STEM scores should *not* correlate strongly (in excess of $r = .50$, see Cohen, 1988) with any dimension of the five factor model. However, given existing relationships between performance-based EI and the five factor model of personality, small to moderate correlations of the STEU and STEM are expected for Agreeableness and Openness to Experience. The Agreeableness link may be stronger for the STEM than the STEU, as higher correlations have been found for Management than Understanding (see Chapter 1, Section 6.2.2). However, this might again be due to method factors, where the stronger relationship of Management to Agreeableness is due to Management's ratings-based response format rather than be construct-related. The two forms of the STEM allow a test of this idea: If Management relates more strongly than Understanding regardless of response format, then both versions of the STEM should relate equally strongly to Agreeableness.

1.2.7 Predictive Validity: Relationships with University Performance

Several researchers have linked EI with academic achievement, based on the idea that social and emotional demands of academic life as well as learning and reasoning with scholastic material are involved in such achievement (Barchard, 2003; Newsome et al., 2000; Parker et al., 2004; Petrides et al., 2004; van der Zee et al., 2002; Woitaszewski & Aalsma, 2004). Despite this finding, prediction of academic achievement from MSCEIT scores has so far been modest: Only the Understanding has related to academic achievement, and this only before intelligence and personality are factored out (Barchard, 2003; Woitaszewski & Aalsma, 2004). However, it is worth re-assessing this relationship with multiple-choice tests, given that Understanding was the only branch related to academic achievement, and is also the only MSCEIT tests in multiple-choice format. It is possible that multiple-choice Management tests may also relate to academic achievement.

In the current study, the incremental predictive validities of the STEM and the STEU are examined for two measures of academic achievement: (1) weighted average mark over all university subjects at the end of the first year of study (WAM); and (2) psychology grade for the first semester of study. It is expected that the STEU and the STEM will predict some variation in University grades, and previous studies might lead us to expect this relationship to be stronger for STEU scores (understanding emotions) than STEM scores (managing emotions).

1.2.8 Assessing EI-Criterion Relationships with Regression: Importance of Incremental Validity

Brody (2004) suggests that the predictive validity of EI should always be evaluated *after* controlling for intelligence and the five factors of personality. So far, the few studies examining the incremental predictive power of EI to important outcomes (after controlling intelligence, personality or both) is weak, indicating that the construct may not be useful (e.g., Austin, Saklofske, & Egan, 2005; Bastian et al., 2005 press; Ciarrochi, Chan, & Bajgar, 2001). In the current study, relationships of STEU and STEM to criterion variables will also be examined *after* controlling for vocabulary and personality test scores to assess the incremental validity of the STEU and STEM.

1.3 Summary of Hypotheses

Hypothesis 1. There will be individual differences in the use of the rating scale, with tendency towards high or low ratings associated with existing individual differences or group membership.

Hypothesis 2. There will be a "method effect" of response format, resulting in a stronger relationship between STEU and STEM when both are in multiple-choice format than when the STEM is ratings-based.

Hypothesis 3. The STEU and STEM will meet Mayer et al.'s (2000b) correlational criteria for an intelligence. That is, STEU, STEM, and Stories scores will converge, with stronger correlations between STEU and STEM as they both measure Strategic EI. STEU and STEM will have moderate but lower relations to vocabulary test scores and the relationships between STEU, STEM, and Stories will remain significant when controlling for vocabulary.

Hypothesis 4. STEU and STEM scores will relate meaningfully to criterion variables. Small negative correlations with alexithymia facets and GHQ scores, and small positive correlations with life satisfaction and academic achievement are expected. These relationships are further expected to hold after personality and vocabulary are controlled.

Hypothesis 5. STEU and STEM scores will not correlate in excess of $r = .50$ with any dimension of the five-factor model of personality, but may show a small degree of relationship to Agreeableness, and Openness to Experience.

2. Method

2.1 Participants

207 Sydney University psychology undergraduate students (140 female) took part in the experiment for course credit. Their median age was 19 ($M = 21.1$, $SD = 5.6$). 55 students were from a rural campus of Sydney University, and the remaining 149 from the main city campus. 110 participants reported their cultural identification as primarily Australian or Anglo-Celtic (e.g., British, American), 77 as primarily from an Asian country, 10 from a Middle-Eastern country, 4 from Europe (non-UK), 3 as Indigenous Australian, and 1 from South America.[15] Cultural identification was coded as either majority (Australian or Anglo-Celtic) or minority (anything else) in further analysis.

2.2 Design

This study used a quasi-experimental design, where two groups of participants completed two different test batteries: "A" or "B". In Battery A the STEM was administered as a multiple-choice test whereas in Battery B the STEM was administered as a ratings-based test. In addition, although Battery A included all tests (in the order listed), Battery B did not include the GHQ-20 (as the ratings-based STEM took approximately 10 minutes longer and time was limited). 113 participants completed Battery A and 94 participants completed Battery B. All participants completing Battery A

[15] 2 participants did not report their gender, 3 participants did not report their age, and 2 participants did not report their ethnicity.

were from the city campus. All participants from the rural campus completed battery B (i.e., 55 rural and 39 city campus undergraduates did Battery B).[16]

2.3. Test Battery

2.3.1 Demographic questions

Before completing the tests listed below, participants answered some brief demographic questions. Participants reported their age, sex, number of years they had been speaking English, the culture or nationality with which they identified.

2.3.2 STEU (Situational Test of Emotional Understanding)

42 multiple-choice questions of the following type were presented to participants: *Xavier completes a difficult task on time and under budget. Xavier is most likely to feel? (a) surprise, (b) pride, (c) relief, (d) hope, (e) joy*. For each of 14 emotions, there were three structurally equivalent items: one that described a situation in a work context, one that described a situation in the context of personal life, and one that described the abstract features of the situation and was thus context-less. The context-less analog for the item presented above was: *By their own actions, a person reaches a goal they wanted to reach. The person is most likely to feel? (a) surprise, (b) pride, (c) relief, (d) hope, (e) joy*). Development of the STEU is described in Chapter 2.

2.3.3 STEM (Situational Test of Emotion Management).

44 questions were presented to participants in either multiple-choice format (in Battery A) or ratings-based format (in Battery B), as shown in the example items below.

---

[16] Although giving all rural students Battery B violated some assumptions relating to random assignment to groups, there were compelling practical reasons for this allocation. Protocol for Battery A originally called for participants to take home two paper-based forms of a relationship quality questionnaire, one of which they got a significant other to fill out, and the other which they filled out themselves. Sending and returning these forms for students at the remote campus was impractical, and for this reason, all students were placed in group B, which did not have this protocol. The relationship-quality questionnaire was not included in the analyses, as the return rate was a little low (75%), and the questionnaire did not appear to be valid (did not significantly correlate with life satisfaction, personality or any other variable in the analysis) so it was therefore excluded from the study.

Multiple-choice STEM item

Clayton has been overseas for a long time and returns to visit his family. So much has changed that Clayton feels left out. What action would be the most effective for Clayton?
(a) Nothing, it will sort itself out soon enough
(b) Tell his family he feels left out
(c) Spend time listening and getting involved again
(d) Reflect that relationships can change over time

Ratings-based judgment STEM item

Clayton has been overseas for a long time and returns to visit his family. So much has changed that Clayton feels left out. How effective are each of the following actions for Clayton?

	Not at all effective	Not very effective	Slightly effective	Somewhat effective	Fairly effective	Extremely Effective
(a) Nothing, it will sort itself out soon enough	1	2	3	4	5	6
(b) Tell his family he feels left out	1	2	3	4	5	6
(c) Spend time listening and getting involved again	1	2	3	4	5	6
(d) Reflect that relationships can change over time	1	2	3	4	5	6

2.3.4 Toronto Alexithymia Scale (TAS-20)

Participants rated 20 self-report items on a five-point scale from Strongly Disagree (1) to Strongly Agree (5). The TAS-20 indexes three aspects of alexithymia: (1) difficulty identifying feelings (DIF; e.g., *I have feelings that I cannot quite identify.*), (2) difficulty describing feelings (DDF; e.g., *People tell me to describe my feelings more*.), and (3) an externally-oriented thinking style (EOT; e.g*., I find examination of my feelings useful in solving personal problems*.). Coffey, Berenbaum, and Kerns (2003) reported Cronbach alpha reliabilities of .79, .70 and .65 for the three sub-scales, respectively.

2.3.5 Satisfaction with Life Scale (SWLS)

Participants rated 5 items of the following type on a seven-point scale from Strongly Disagree (1) to Strongly Agree (7): *In most ways my life is close to ideal*. Pavot and Diener (1993) provide comprehensive information on the reliability and validity of the SWLS. Reported alpha reliability ranges from 0.79 to 0.89 and test-retest reliability ranges from $r = 0.50$ (over 10 weeks) to $r = 0.84$ (over 4 weeks). Construct validity of the SWLS has been demonstrated through group differences (between "normals" and psychiatric patients, prisoners, students in poor or turbulent counties, and abused women); and high correlations with previous life satisfaction scales, interviewer ratings, and

observer ratings of life satisfaction. In line with previous structural analysis of the SWLS, scores are analysed in terms of two factors: current satisfaction and retrospective satisfaction (McDonald, 1999).

2.3.6 Stories

The Stories test was taken from the Multi-Factor Emotional Intelligence Scale (MEIS, Mayer et al., 1997). Participants were presented with six stories of 2-3 sentences describing events happening to a fictitious person. For each story, participants rated the degree of emotion they thought the protagonist felt for seven different emotions. Ratings were given on a five-point scale from "definitely present" to "definitely NOT present". Caruso, Mayer, and Salovey (2002) reported Cronbach alpha reliabilities of .78 to .85 with test-retest reliability of r = .60 (over 2 weeks). The Stories test has the strongest loading of any test on a general EI factor composed of all twelve MEIS tests (Mayer et al., 2000b; Roberts et al., 2001).

2.3.7 Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed 20-item version (OCEANIC-20)

A 20-item short form of the OCEANIC was constructed for this study by taking four items from each dimension of the existing 60-item OCEANIC (items were selected for their high loadings on each dimension as reported in the OCEANIC manual, Roberts et al., 2000). Although such short forms may have somewhat lower reliabilities than comprehensive measures, they are often used in research situations for their efficiency, particularly when personality is not a major focus of the study (see Gosling, Rentfrow, Swan, 2003). Participants rated the 20 items on a five-point scale from Never (1) to Usually (5)[17]: e.g., *I am talkative*. The OCEANIC manual reports Cronbach alphas of between .80 (for Openness to Experience) to .90 (for Extraversion) for the full 12-item scales (Roberts, 2000).

[17] The OCEANIC normally employs a 6-point scale from Never (1) to Always (6). An unfortunate bug in the computer program meant that the sixth point (Always) was not visible to participants, and therefore the scale was effectively only 5 points, labeled Never, Rarely, Sometimes, Often and Usually.

2.3.8 Vocabulary

This 18-item test follows the format of the Wechsler vocabulary test (WAIS-III, Wechsler, 1997) and was taken from Stankov's (1997) Gf/Gc quickie test battery to represent the primary ability of verbal comprehension and broad ability of crystallized intelligence (Gc). Participants were given a target word and asked to select the word most similar to the target from five alternatives. *Eg: revolve: 1. A gun, 2. Uprising, 3. Turn around, 4. Grow, 5. Decide*. Estimates of Cronbach alpha reliability range from .75 to .79 in an undergraduate sample, and factor analyses based on Gf/Gc theory confirms this as a valid marker of Gc. Several published studies have included this test as a Gc marker (e.g., Davies et al., 1998; MacCann et al., 2004; Pallier, 2003).

2.3.9 General Health Questionnaire (GHQ-20)

Participants rated 20 psychological health-related statements such as "Have you recently lost much sleep over worry" on a four-point scale where higher ratings indicated increasing levels of distress. Scale points were scored from 0 to 3, according to established procedure, resulting in a total score that could theoretically range from 0 to 60 (Goldberg, 1972). Participants were asked to consider their general health over the past few weeks only in answering these 20 items. The original 60-item scale of the GHQ was created to detect psychiatric illness of non-psychotic types (Goldberg, 1972). The shorter GHQ-20 has been shown to discriminate between psychiatric patients and non-patients (Chan, 1995) and has been used as a measure of psychological distress (Khavarpour & Rissel, 1997; Snekkevik, Anke, Stanghelle, & Fugl-Meyer, 2004).

2.4 Procedure

Participants read through an information form and then completed a consent form indicating their willingness to participate and whether or not they were willing for the experimenter to access and use their university grades (128 psychology grades and 117 weighted average grades over all

tertiary subjects were available for analysis).[18] The 152 participants from the city campus read paper and pencil information and consent forms, and completed the test battery outlined below on PC computers in groups of approximately 10 in on-campus labs. The 55 participants from the rural campus were emailed information, consent forms, and a computerised version of the test battery, which they completed at computer terminals of their choosing. Although all tests were self-paced, testing took approximately 1 hour. This study was approved by the Sydney University Human Research Ethics Committee.

2.5 Analyses

There were several stages of analyses in this study. For the purposes of clarity, these are outlined below.

Analysis 1: Testing for Sampling Error. This analysis checked whether there were any differences between participants completing Battery A and Battery B. Reliabilities, means, and standard deviations over all scores were compared separately for those completing Battery A (Group A) versus Battery B (Group B).

Analysis 2: Testing for Individual Differences in Use of Ratings Scales for EI Measures. This analysis examined whether participants' raw ratings on the Stories or ratings-based STEM related meaningfully to personality traits, gender and self-identified culture. Differences in use of the rating scale for experts and non-experts were also examined.

Analysis 3: Zero-order and Partial Correlations between EI and Vocabulary. Zero-order correlations between EI and vocabulary were calculated to provide evidence of convergent validity of the STEU and STEM (i.e., whether they related to each other and a Stories and could be distinguished

[18] For many rural students, grades could not be collected, as they gave their "local" student identification number rather than the official one on the "permission to access grades" consent form, which could not be matched to the central records of grades. For this reason, the relationship of the ratings-based STEM to academic achievement could not be assessed in the current study.

from vocabulary). In addition, partial correlations between EI measures controlling for vocabulary were examined to determine if shared variance among EI measures was due to verbal ability.

Analysis 4: Regressions Predicting Criterion Measures from EI. Two sets of regressions predicting criterion measures from STEU and STEM scores were undertaken: (1) simple linear regression, where *only* the EI score predicted the criterion; and (2) three-step hierarchical regression, where vocabulary was entered as the first step, the five factors of personality entered as the second step, and the EI variable entered as the third step. Comparing the squared multiple correlations (R^2s) for the simple linear regressions to those from the third step of the hierarchical regressions (ΔR^2s) shows whether relationships between criteria and EI scores still hold after vocabulary and personality scores are controlled. This analysis was undertaken to provide incremental validity evidence for the STEU and STEM.

Analysis 5: Structural Examinations of the Distinction between Understanding and Management. Subsets of the STEU and STEM were used as indicators in structural equation modelling. Nested models compared a one-factor model (where all subsets defined one Strategic EI factor) to a two-factor model (where STEU subsets defined an Understanding factor and STEM subsets defined a Management factor) to determine whether the two-factor model provided incrementally better fit to the data. This analysis was undertaken twice: once for Battery A scores (i.e., the multiple-choice STEM) and once for Battery B scores (i.e., the ratings-based STEM). This purpose of this analysis was to determine whether Understanding and Management were more strongly related when both were measured by multiple-choice tests.

3. Results

3.1 Testing for Sampling Error

Group A (who completed the STEM as multiple-choice) and Group B (who completed the STEM as ratings-based) had significantly different scores on two EI variables, outlined in Table 3.1. Group A's performance on the STEU and the externally-oriented thinking facet of the TAS-20 was significantly better than Group B's performance. Test scores are also less reliable for Group A

(perhaps indicating a restriction of range effect). These sampling error differences are problematic, as differences between the multiple-choice STEM and ratings-based STEM may be due to sample differences, rather than differences in item format. Given that Group B's test scores are more reliable than Group A's on some criterion measures, the ratings based STEM may be more reliable than the multiple-choice STEM *even if rating scales have no effect on reliability*. Also, correlations between these criterion variables and STEM scores might be higher for ratings-based STEM because the criterion variables are more reliable (rather than the ratings-based format being more valid). Further results must be interpreted with these sampling differences in mind.

Table 3.1

Differences in Means, Standard Deviations, and Cronbach Alpha Reliability for All Variables Showing Significant Differences between Experimental Groups

	Group A (N = 113)			*Group B (N = 94)*			*Compare Means*	*Compare Variation*
	Mean	SD	α	Mean	SD	α	*t*	*F*
STEU (standards)	0.62	0.11	.64	0.57	0.13	.75	2.943**	4.698*
EOT	16.37	4.45	.71	19.35	4.81	.68	4.545**	0.081

Note. Group A completed the STEM in MC format, group B in ratings-scale format. EOT = Externally-Oriented Thinking. There were no significant differences between group means on any other variable reported in this study.

*$p < .05$, **$p < .01$

3.2 Individual Differences in the Use of the Rating Scale

Before scoring the test, participants' spread of raw ratings from 1 to 5 across both the STEM and the Stories were examined. Each participant's mean rating, standard deviation of ratings, and three quartile scores (25th, median, and 75th) over all test items, were calculated for both measures. The mean and median showed what part of the scale participants tended to stick to (e.g., if the mean and median were high, participants would tend to rate all options on the STEM as effective and all

options on the Stories as definitely present). The standard deviation would represent how much they spread out their answers along the whole scale, instead of using only one region or part of the scale.

Table 3.2 shows group differences in these parameters by sex and by majority-culture identification, as well as the relationship between these parameters and personality dimensions. The effect sizes for the difference between this sample and the experts who rated the STEM are also shown (see Study 3 of Chapter 2 for a description of expert ratings). If meaningful group differences or correlations with other variables exist, then individual differences in rating-scale use are non-random and meaningful. Of the 50 correlations, only 2 or 3 would be significant at $p < .05$ (and 0 or 1 at $p < .01$) if differences in rating-scale use were not meaningful, whereas 7 of these show significant relationships at $p < .05$, and 2 at $p < .01$.

Although differences between experts and non-experts were not significant, effect sizes were quite large for mean and median ratings, with experts tending to rate actions as *less* effective than novices (as there were only 5 experts, and 92 non-experts, statistical significance would require a very large effect size). There are no gender differences in rating-scale use for the Stories test, in agreement with Mayer et al. (2000b). However, there *are* gender differences in rating scale use on the STEM, where females tended to give higher effectiveness ratings. Participants from the majority culture (Australian or Anglo-Celtic) had less variation in Stories ratings and higher minimum ratings than others. Conscientiousness showed no relationship with rating parameters on the Stories test, but related to greater variation in STEM ratings (higher SDs and higher ratings at the 75^{th} percentile). High ratings on the STEM relate to Agreeableness (possibly indicating a social-desirability effect or demand characteristic).

Table 3.2

Correlations between the Mean, SD, and Quartile Cut-Offs for Raw Ratings of the STEM and Stories, with Personality, with Group Differences by Expertise, Gender and Culture

	STEM Ratings					*Stories Ratings*				
	Mean	*SD*	*P25*	*P50*	*P75*	*Mean*	*SD*	*P25*	*P50*	*P75*
Correlations										
Openness	.20	-.10	.09	.20	.14	.06	-.08	.12	.03	-.01
Conscientiousness	.13	.25*	-.14	.13	.29**	.04	-.01	.11	-.04	.08
Extraversion	.02	.03	.04	.01	.06	.04	.01	.05	.05	.07
Agreeableness	.29**	.22*	.01	.27*	.25*	.08	.08	.01	.07	.16*
Neuroticism	.03	.06	-.05	-.11	.04	.06	.07	-.01	.05	.04
Difference Scores										
Expertise	-0.67	0.11	-0.69	-0.27	-0.18	-	-	-	-	-
Gender	-0.36	-0.21	-0.27	-0.49*	-0.47*	0.00	0.09	-0.10	0.12	-0.08
Culture	-0.03	-0.13	0.07	-0.03	-0.16	0.17	-0.37*	0.45**	0.14	-0.09

Note. Difference scores are calculated as expert – non-expert (i.e., negative values indicate higher ratings by non-experts), male – female (i.e., negative values indicate higher ratings given by females), and majority culture – non-majority culture (i.e., negative values indicate higher ratings given by cultural minorities)[19]; significance of the difference scores refers to *t*-values.
*$p < .05$, **$p < .01$

Generally, Table 3.2 shows that there *are* meaningful individual differences in the way individuals use the rating scale for the STEM and Stories tests. Evidence of individual differences in the use of the rating scale indicates that profile scoring (or scoring which standardises individuals'

[19] The formula used to calculate effect size in this and subsequent chapters is the difference in means divided by the pooled standard deviation (where the pooled SD accounted for different sample sizes in the two different groups). That is, the formula is: $(\mu_1 - \mu_2)/\sqrt{((\sigma^2_1(n_1-1)+(\sigma^2_2(n_2-1))/(n_1+n_2-2)}$

ratings) may provide a more valid measure of the construct than distance-based scores. For this reason, profile scores are used in further analyses of the Stories and the ratings-based STEM.

3.3 Reliability and Descriptive Statistics

3.3.1 Reliability and Descriptive Statistics for EI Variables

Descriptive statistics for the EI measures are shown in Table 3.3.[20] Although females generally did better than males on EI tests (as previous research would suggest; Ciarrochi et al, 2000; Day & Carroll, 2004; Kafetsios, 2004; Mayer, et al., 2000b), these mean differences were not significant. Reliabilities are acceptable for group assessment for the STEU and STEM (except for the standards-scored version) but not the Stories (i.e., greater than .60; Wasserman & Bracken, 2003). Using the coping-derived scoring-key to score the STEM resulted in scores that were virtually random error, and is clearly not a useful scoring system. Given that the reliability is effectively zero, no other analyses of the standards-scored STEM will be undertaken.

Screening for non-normality indicated no problems with non-normal distributions shapes (i.e., skew < 2, kurtosis < 7 for all variables; West, Finch, & Curran, 1995). For both the ratings-based EI tests (STEM and Stories), distance-based scoring was more reliable than proportion-based scoring, and there was little difference in reliability between profile and non-profile scores (i.e., distance and standardised distance). In further analyses, only the profile scores of the ratings-based STEM and Stories are used. In addition, item 4 of Stories (which had a low item-total correlation) was removed, increasing reliability to .43 (M = -2.84, SD = .43 for Stories after this item was cut). The mean expert

[20] Several different methods of scoring each test were under-taken, although not all are reported for reasons of efficiency. Different types of profile scores (Pearson, Spearman and Standardised) were highly correlated for both consensus and expert scores for both the STEM and Stories (r = .93 to r = .99). For this reason, only standardised scores (and not Pearson or Spearman scores) were included in Table 3.3 (standardised scores standardise over all ratings, whereas correlational scoring standardises within each item). Consensus scores for the both versions of the STEM, as well as the STEU and Stories were calculated, and generally had higher skew and kurtosis than expert or veridical scores, and were slightly more reliable. For distance-based consensus scores, Mahalanobis distance scores were also calculated, and were no different from Euclidean distance scores for the STEU (r = .98) and slightly different but less reliable for Stories (r = .82; α = .48).

rating version of the multiple-choice STEM was used in further analyses as this was slightly more reliable.

Table 3.3

Descriptive Statistics and Reliabilities for EI (Comparing Different Response Formats and Scoring Types)

	N	*Alpha* [a]	*Mean (SD)*			*Gender d* [b]
			All	*Male*	*Female*	
STEM (multiple-choice)						
1a. Standards	112	.09	0.53 (.06)	0.53 (.06)	0.52 (.05)	0.11
1b. Mean rating [c]	112	.74	0.75 (.03)	0.74 (.04)	0.75 (.02)	-0.35
1c. Proportions	112	.68	0.52 (.07)	0.50 (.09)	0.53 (.06)	-0.35
STEM (ratings-based)						
2a. Distance	91	.92	2.57 (.46)	2.54 (.50)	2.58 (.45)	-0.10
2b. Proportions	91	.86 [.86]	0.28 (.05)	0.27 (.05)	0.28 (.05)	-0.16
2c. Standardised distance [c]	91	.91	3.57 (.28)	3.50 (.30)	3.60 (.27)	-0.34
STEU						
3. Standards	200	.71	0.60 (.13)	0.59 (.13)	0.60 (.12)	-0.11
Stories						
4a. Distance	201	.52	1.91 (.47)	1.88 (.51)	1.92 (.45)	-0.10
4b. Proportions	201	.37 [.53]	0.78 (.08)	0.77 (.09)	0.78 (.08)	-0.15
4c. Standardised distance [c]	201	.38	-2.41 (.33)	-2.46 (.33)	-2.38 (.33)	-0.23

[a]For ratings-based tests scored by proportions, multiple ratings of each stimulus were averaged before calculating reliability to avoid reliability inflation due to experimental dependence. Cronbach alphas considering all ratings as separate items are shown in brackets, to allow comparison with the MEIS and MSCEIT, where alphas are calculated this way. [b]Effect sizes for gender differences are Male – Female, such that negative values indicate that females receive higher scores. [c]These scores are used in further analysis.

Different methods of scoring converged reasonably well, indicating reasonable consistency among scoring methods. Mean-rating and proportion scores on the multiple-choice STEM were highly correlated (r = .92), as were distance and proportion scores on Stories and the ratings-based STEM (r = .95 and r = .92, respectively). However, standardised distance scores were less strongly related to (non-standardised) distance and proportion scores for both the STEM (r = .61 and r = .63 respectively) and Stories (r = .45 and r = .51 respectively). It appears that standardisation does substantially change scores.

3.3.2 Reliability and Descriptive Statistics for Criterion Variables

Reliability estimates and descriptive statistics for the criterion variables are shown in Table 3.4. Reliability is greater than .70 for all scores, indicating they are easily reliable enough to use in research. Reliabilities for the 20-item OCEANIC were slightly lower than those reported for the full 60-item version in the manual (Roberts, 2000), but never by more than .10, which seems reasonable for a two-thirds reduction in the number of items. Descriptive statistics and reliability estimates were in line with prior findings for all measures except for the General Health Questionnaire (GHQ). A one sample t-test showed the current sample was significantly *more* distressed than Jacobsen, Hasvold, Hoyer, and Hansen's (1995) sample (t_{109} = 2.682, p = .008). Indeed, a large proportion of this sample (38.2%) scored above Goldberg's (1972) reported GHQ mean among people with severe mental illness. Even given that Goldberg's sample are British and that their scores were obtained more than 30 years ago, the size of this effect seems to indicate that the GHQ is not functioning as a measure of general health in the current study. It seems far more likely that demand characteristics are responsible for this result than that half the sample is severely distressed (particularly since their life satisfaction is at normal levels). Since participants completed the GHQ after the OCEANIC, TAS-20, and SWLS, they may have been primed to consider the GHQ a personality questionnaire rather than a symptom checklist and thus attempted to vary their answers along the scale as in a personality test, rather than report that they had no symptoms in the majority of items. GHQ-20 scores might thus not be functioning well in this study, and results for the GHQ will be interpreted with caution.

Table 3.4

Descriptive Statistics and Reliabilities for EI (Comparing Different Response Formats and Scoring Types)

	Current Study					*Comparison Group*	
	N	*Alpha*	*Mean*	*SD*	*Gender d* [a]	*Alpha*	*d*
5. Vocabulary [b]	201	.79	0.49	0.20	0.00	.75	.01
6. Openness [c]	199	.71	13.21	3.05	0.39**	.80	-
7. Conscientiousness	199	.82	13.99	3.19	0.04	.87	-
8. Extraversion	199	.85	12.42	3.56	-0.32*	.90	-
9. Agreeableness	199	.77	16.97	2.39	-0.14	.88	-
10. Neuroticism	199	.75	12.71	3.15	-0.36*	.87	-
11. Difficulty Identifying Feelings (DIF) [d]	200	.81	16.87	5.52	-0.35*	.85	.01
12. Difficulty Describing Feelings (DDF)	200	.75	13.56	4.32	-0.10	.84	.05
13. Externally-Oriented Thinking (EOT)	200	.73	17.68	4.83	0.41**	.63	.01
14. Life Satisfaction (SWLS) [e]	201	.85	23.31	6.23	0.01	.87	-.03
15. SWLS-Retrospective	201	-	23.31	6.23	0.09	-	-
16. SWLS-Current	201	-	14.49	3.89	-0.09	-	-
17. General Health Questionnaire (GHQ-20) [f]	110	.91	19.84	9.53	0.30	.90	.32**
18. University Admissions Index (UAI)	161	-	88.27	11.98	-0.30	-	-
19. Psychology 1 grade	128	-	68.03	11.67	-0.26	-	-
20. Weighted Average Mark (WAM)	117	-	66.76	13.40	0.02	-	-

[a]Effect sizes for gender differences are Male – Female, such that negative values indicate higher scores for females. [b]Vocabulary comparison from MacCann et al.'s (2004) data. [c]OCEANIC comparison reliabilities from Roberts (2000); note that *d*-scores were not calculated since this study used a short form of the OCEANIC). [d]TAS-20 comparison from Ciarrochi et al. (2003). [e]SWLS comparison from Diener et al. (1985). [f]GHQ comparison from Jacobsen et al. (1995) for descriptive statistics, and Goldberg (1972) for reliability.

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

Males were significantly more Open to Experience, less Extraverted, and less Neurotic than females, consistent with prior findings (Costa, Terracciano, & McCrae, 2001; Feingold, 1994).[21] Gender differences for the TAS-20 were also consistent with previous findings, with females scoring significantly more highly on Difficulty Identifying Feelings, and males on Externally-Oriented Thinking (e.g., Hexel, 2003; Montebarocci, Codispoti, Baldaro, & Rossi, 2004).

3.3 Correlations of EI Measures with Vocabulary and Personality

Table 3.5 shows the Pearson correlations between the EI tests (STEU, STEM, and Stories) with and without controlling for vocabulary, as well as their zero-order correlations with vocabulary and personality. As hypothesised, STEU and STEM scores are highly correlated with each other, although the relationship is stronger for the multiple-choice than the ratings-based STEM. The STEU and ratings-based STEM are moderately related to the Stories test as hypothesised, but the multiple-choice STEM is unrelated to Stories. That is, multiple-choice tests correlate more highly with multiple-choice tests and ratings-based tests correlate more highly with ratings based tests, supporting the hypothesis that response-format acts as a method effect.

The correlation between multiple-choice STEU and STEM is extremely high when correcting for reliability (r = .90). It appears that Understanding and Management may not be differentiable when both are in multiple-choice format (rather than when Understanding is multiple-choice and Management is ratings-based, as is the case in the MSCEIT). This issue is explored further through structural analysis in Section 5 of these results.

[21] Males tend to score higher on openness to ideas (the intellect-related facets of Openness) whereas woman tend to score higher on openness to aesthetics and feelings (the emotion-related facet of Openness). This study did not distinguish between these two types of Openness, but all four Openness questions related more to intellectual Openness than to Openness to Feelings (items were “I am philosophical”, “I think about the wonders of nature”, “I think about the origins of the universe” and “I am inventive”).

Table 3.5

Correlations between EI and Criterion Variables (Partial Correlations between EI Scores with Vocabulary, Gender, and Years of English Speaking Variation Removed are Italicised in Top Right of Matrix)

	STEU	STEM (MC)	STEM (Ratings)	Stories
STEU	--	*.55***	*.42***	*.32***
STEM (MC)	.65***	--	--	*-.13*
STEM (Ratings)	.49***	--	--	*.42***
Stories	.34***	-.04	.44***	
Vocabulary	.49***	.40***	.26*	.13
Openness to Experience	.11	-.15	.00	-.02
Conscientiousness	-.02	-.01	.22*	.04
Extraversion	-.06	.08	.03	-.10
Agreeableness	.16*	.23*	.31**	-.10
Neuroticism	-.04	-.02	-.09	.01

Note. The multiple-choice STEM is scored by mean expert rating weights; ratings-based STEM by standardised distance; STEU by the standards-based key and Stories by standardised distance.

For the STEU and ratings-based STEM, relationships with vocabulary are similar to those reported for the MEIS and MSCEIT Understanding and Management. However, correlations with vocabulary were somewhat higher for the multiple-choice STEM than reported for the MSCEIT Management tests, where Understanding scores relate more strongly to intelligence than Management scores (Mayer et al., 2000b, found $r = .40$ for Understanding, $r = .20$ for Management; Lopes et al., 2003, found $r = .39$ for Understanding, and no significant correlations with other branches). The STEU correlated more highly with vocabulary than the ratings-based STEM ($z = 2.11$, $p = .035$), but not the multiple-choice STEM ($z = 0.94$. $p = .347$). MSCEIT-based findings that Understanding correlates more strongly with intelligence than the other branches may be due to response format issues. As expected, the multiple-choice STEM correlated more strongly with vocabulary than the

ratings-based STEM, but not significantly so (z = 1.100 p = .136). The expectation that the ratings-based STEM would relate more strongly to Agreeableness than the multiple-choice STEM was also met, although again this also was not significant (z = 0.600, p = .274).

In addition, both the STEU and multiple-choice STEM were more strongly related to vocabulary (an established intelligence measure) than they were Stories (an established EI measure). Although Stories had much lower reliability than vocabulary, STEU and multiple-choice STEM scores were still slightly more strongly related to vocabulary than Stories even when correcting for reliability (r = .62 and -.07 for Stories and r = .65 and .52 for vocabulary for the STEU and STEM, respectively). That is, EI tests relate more strongly to intelligence than to other EI tests – a poor indication of discriminant validity. Such a result does not support Mayer et al.'s (2001) correlational criteria of shared and unique variance among EI tests needed for EI to attain the status of an intelligence.

Performance-based EI was clearly independent from personality, with the highest correlation at r = .31 (Cohen's [1988] description of "small" is r = .30). As hypothesised, all three EI measures related to Agreeableness, although the expected relationship with Openness to experience was not observed. This relationship with Agreeableness is consistent with similar findings for the MSCEIT for Management, but not for Understanding. Day and Carroll (2004), Lopes et al. (2004), Lopes et al. (2003), and Warwick and Nettelback (2004) found small correlations between Agreeableness and Management but *not* between Agreeableness and Understanding. In this study, the relationship to Agreeableness was not significantly different for STEU and STEM, for either the multiple-choice STEM (z = -.61, p = .542) or the ratings-based STEM (z = -1.24, p = .215).

Correlations between the STEU, STEM, and Stories were similar after vocabulary, gender and English speaking were accounted for, decreasing only slightly. That relationships among these three text-based EI measures are *not* merely due to verbal skill provides some support for the unique variation criterion for EI as a standard intelligence (although stronger relationships between multiple-choice EI measures and vocabulary than such measures and Stories do not support this criterion). In

general, correlations indicate that the STEU and STEM are high enough to indicate that all measure the same construct, although may overlap with vocabulary. They are largely independent from personality, strongly relate to each other, and moderately relate to Stories even when vocabulary is accounted for.

3.4 Relationships of the STEU and STEM to Criterion Variables

The incremental prediction of criterion variables (alexithymia, life satisfaction, emotional distress, and academic achievement) from STEU and STEM scores was assessed via a series of three-step hierarchical regression models.[22] In the first step, only the vocabulary score was entered. In the second step, the five factors of personality were added (i.e., predictors were vocabulary, Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism). In the third step, the EI score (STEU, multiple-choice STEM, or ratings-based STEM) was added to the model. The change in R^2 (ΔR^2) for each step is given in Tables 3.6, 3.7 and 3.8, showing whether EI scores predict criteria over and above the effect of vocabulary and personality scores. In addition, these tables show the R^2 values for simple linear regressions with EI as the only predictor (i.e., the squared zero-order correlation between the EI score and each criterion measure). This allows a test of whether EI predicts criteria at all, even before vocabulary and personality are accounted for. The parameter values for each regression equation are given in Appendix 3.3 rather than in the tables, for reasons of space and efficiency.

3.4.1 Relationships with Alexithymia

The STEU and STEM's prediction of alexithymia is shown in Table 3.6. After controlling for vocabulary and personality, the STEU, multiple-choice STEM and ratings-based STEM predicted 6%, 7% and 5% of the variation in externally-oriented thinking respectively. Although such

---

[22] The correlations between criterion variables are given in Appendix 3.1. As some of the EI and criterion variables showed significant gender differences, correlations were calculated separately for males and females, but there did not appear to be any difference in correlation by gender. In addition, variables were mean-corrected separately by sex, and correlation matrices were calculated for these mean corrected variables. These are given in Appendix 3.2.

increments are relatively small, results illustrate that the STEU and STEM are capturing some incremental variance in an emotion-related criteria. The other two components of alexithymia had no significant relationships to either the STEU or STEM even before controlling for vocabulary or personality.

3.4.2 Relationships with Life Satisfaction and Emotional Distress

The STEU and STEM's prediction of life satisfaction (both current levels and retrospective satisfaction) and General Health Questionnaire (GHQ) scores are shown in Table 3.7. As hypothesised, the multiple-choice STEM predicted retrospective (but not current) levels of life satisfaction even when personality and intelligence were accounted for, accounting for an additional 4% of the variance in scores. GHQ scores were not significantly predicted by either STEU or STEM scores but were strongly predicted by personality factors. This is consistent with the idea that participants were answering the GHQ as if it were a personality measure rather than a checklist of symptoms.

Table 3.6

Squared Multiple Correlations (R^2s) and Incremental R^2s (ΔR^2s) for Three-Step Hierarchical Regression Models Predicting Alexithymia from Vocabulary in Step 1, Personality in Step 2, and EI in Step 3(R^2 for Simple Linear Regressions Using Only EI Also Shown)

	TAS-DIF			TAS-DDF			TAS-EOT		
	Step 1 (Voc)	Step 2 (FFM)	Step 3 (EI)	Step 1 (Voc)	Step 2 (FFM)	Step 3 (EI)	Step 1 (Voc)	Step 2 (FFM)	Step 3 (EI)
Set 1: Regressions Using STEU scores (N = 189)									
R^2	.01	.33**	.33**	.00	.30**	.30**	.06**	.18**	.24**
ΔR^2	.01*	.32**	.00	.00	.29**	.00	.06**	.12**	.06**
R^2 *(EI only)*[a]			.01			.02			.14**
Set 2: Regressions Using Multiple-Choice STEM scores (N = 109)									
R^2	.02	.35**	.36**	.01	.32**	.33**	.06*	.21**	.28**
ΔR^2	.02	.33**	.01	.01	.32**	.01	.06*	.16**	.07**
R^2 *(EI only)* [a]			.00			.00			.13**
Set 3: Regressions Using Ratings-Based STEM Scores (N = 82)									
R^2	.01	.41**	.41**	.00	.26**	.26**	.06*	.23**	.28**
ΔR^2	.01	.40**	.00	.00	.26**	.00	.06*	.18**	.05*
R^2 *(EI only)* [a]			.03			.01			.11**

Note. TAS-DIF = Difficulty Identifying Feelings, TAS-DDF = Difficulty Describing Feelings, TAS-EOT = Externally Oriented Thinking, Voc = Vocabulary Score, FFM = Five factor model personality dimensions, EI = Emotional Intelligence Score (STEU in set 1, multiple-choice STEM in set 2 and ratings-based STEM in set 3).

[a] R^2 *(EI only)* refers to the squared zero-order correlation between the criterion variables and EI (i.e., vocabulary and personality are not accounted for)

*$p < .05$, **$p < .01$.

Table 3.7

Squared Multiple Correlations (R^2s) and Incremental R^2s (ΔR^2s) for Three-Step Hierarchical Regression Models Predicting Life Satisfaction and GHQ scores from Vocabulary in Step 1, Personality in Step 2, and EI in Step 3(R^2 for Simple Linear Regressions Using Only EI Also Shown)

	Life Satisfaction (current)			Life Satisfaction (retrospective)			GHQ		
	Step 1 (Voc)	Step 2 (FFM)	Step 3 (EI)	Step 1 (Voc)	Step 2 (FFM)	Step 3 (EI)	Step 1 (Voc)	Step 2 (FFM)	Step 3 (EI)
1. Regressions for STEU scores (N = 189, except for GHQ prediction where N=106)									
R^2	.00	.17**	.18**	.01	.20**	.21**	.05*	.28**	.28**
ΔR^2	.00	.17**	.01	.01	.19**	.01	.05*	.23**	.00
R^2 *(EI only)* [a]			.01			.02			.01
2. Regressions for Multiple-Choice STEM scores (N = 108)									
R^2	.02	.16**	.16*	.04*	.22**	.25**	.05*	.27**	.28**
ΔR^2	.02	.13**	.00	.04*	.18**	.04*	.05*	.22**	.01
R^2 *(EI only)* [a]			.01			.08**			.03
3. Regressions for Ratings-Based STEM Scores (N = 82)									
R^2	.01	.36**	.00	.00	.32**	.01			
ΔR^2			.00			.02			
R^2 *(EI only)* [a]	.01	.36**	.36**	.00	.32**	.33**			

Note. Voc = Vocabulary Score, FFM = Five factor model personality dimensions, EI = Emotional Intelligence Score (STEU in set 1, multiple-choice STEM in set 2 and ratings-based STEM in set 3).

[a] R^2 *(EI only)* refers to the squared zero-order correlation between the criterion variables and EI (i.e., vocabulary and personality are not accounted for)

*$p < .05$, **$p < .01$.

3.4.3 Relationships with Academic Achievement

Table 3.8 shows the prediction of psychology grade and weighted average mark from the STEU and multiple-choice STEM. Without accounting for variation due to vocabulary and personality, both the STEU and multiple-choice STEM predicted psychology grade at the end of the

first semester of University. However, only the STEU related to the overall grade at the end of the first year of university (Weighted Average Mark, or WAM). When personality and vocabulary were controlled, STEU scores no longer predicted students' WAMs but did incrementally predict 8% of the variation in psychology grades.

Table 3.8

Squared Multiple Correlations (R^2s) and Incremental R^2s (ΔR^2s) for Three-Step Hierarchical Regression Models Predicting Academic Achievement from Vocabulary in Step 1, Personality in Step 2, and EI in Step 3 (R^2 for Simple Linear Regressions Using Only EI Also Shown)

	Psychology 1 Grade			*Weighted Average Mark for First Year University*		
	Step 1 (Voc)	*Step 2 (FFM)*	*Step 3 (EI)*	*Step 1 (Voc)*	*Step 2 (FFM)*	*Step 3 (EI)*
1. Regressions for STEU scores (N = 109)						
R^2	.09**	.20**	.28**	.08**	.13*	.16*
ΔR^2	.09**	.11*	.08**	.08**	.05	.03
R^2 *(EI only)* [a]			.18**			.13**
2. Regressions for Multiple-Choice STEM scores (N = 103)						
R^2	.14**	.25**	.27**	.07**	.14*	.14*
ΔR^2	.14**	.10*	.02	.07**	.08	.00
R^2 *(EI only)* [a]			.07**			.01

Note. Voc = Vocabulary Score, FFM = Five factor model personality dimensions, EI = Emotional Intelligence Score (STEU in set 1, multiple-choice STEM in set 2 and ratings-based STEM in set 3).
[a] R^2 *(EI only)* refers to the squared zero-order correlation between the criterion variables and EI (i.e., vocabulary and personality are not accounted for)
*$p < .05$, **$p < .01$.

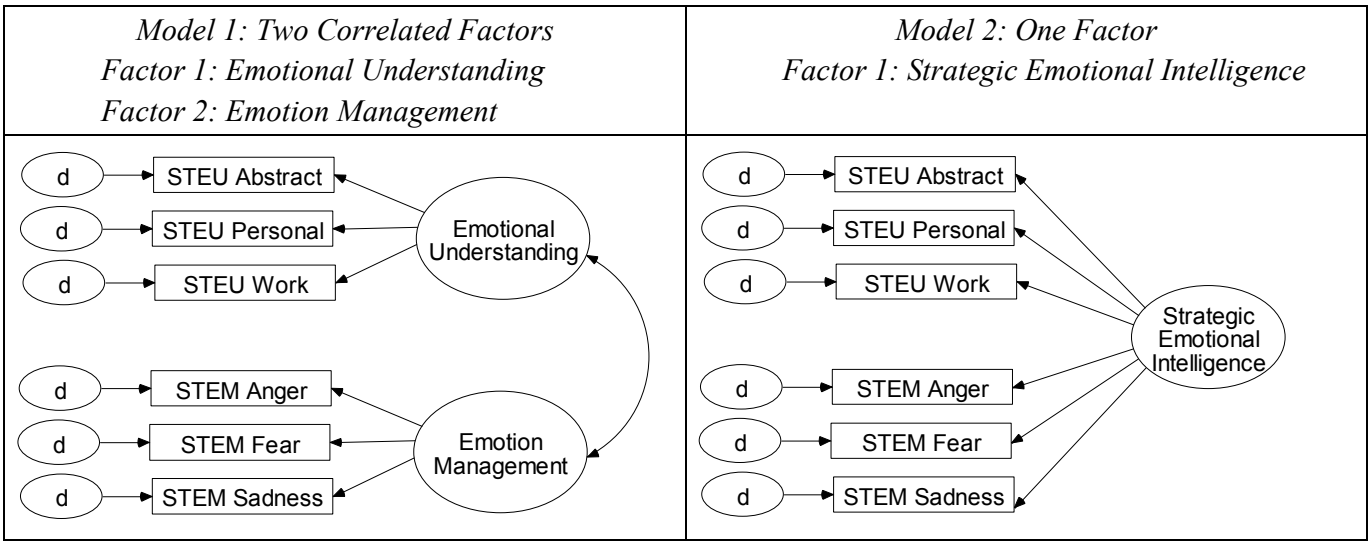
3.5 Nested Models Differentiating Understanding and Managing Emotions

There is more error in multiple-choice than ratings-based STEM scores, making it difficult to compare the relationship of STEM to STEU for multiple-choice versus ratings-based response

formats. Structural equation models examining the relationship between Understanding and Management factors predicted by STEU and STEM subtests explicitly model this error, and thus might be more informative than zero-order correlations. The nested models design shown in Figure 3.1 tests whether a two-factor solution differentiating between Understanding and Management factors fits better than a one-factor solution. The STEU is partitioned into three scores based on whether item content represents a personal-life context (e.g., Penny's hockey team train hard and win the championship), a work life context (e.g., Xavier completes a difficult task on time and under budget) or an abstract context (e.g., By their own actions, a person has reached a goal they wanted to reach). The STEM is partitioned into three scores based on whether item content represents fear-management, sadness-management, or anger-management (see Chapter 2 for a full description of the STEU and STEM's development).

Figure 3.1

Nested Models Design Differentiating Emotional Understanding and Emotion Management



Items that lowered the internal consistency of the STEU and STEM subsets were removed before subsets were calculated. Reliability, descriptive statistics, and subset correlations for the STEU and STEM subsets are shown in Table 3.9 for both the multiple-choice and ratings-based STEM. This analysis was run separately for the multiple-choice and ratings-based STEM, to test whether there is a

different relationship between Management and Understanding based on response format. Missing data were deleted list-wise such that the $N = 108$ for analyses involving the multiple-choice STEM and $N = 88$ for items involving the ratings-based STEM.

Table 3.9

Descriptive Statistics, Reliability and Correlations between Subsets of the STEM and STEU (Parameters for the Multiple-Choice STEM on the Left; Parameters for the Ratings-Based STEM Italicised on the Right)

	Descriptives: MC STEM		*Correlations*						*Descriptives: Ratings STEM*	
	Mean/SD	*α*	*1*	*2*	*3*	*4*	*5*	*6*	*Mean/SD*	*α*
STEU										
1. Abstract (8 items)	.68 (.20)	.47		*.57*	*.57*	*.40*	*.34*	*.39*	--	--
2. Personal Life (10 items)	.66 (.21)	.55	.40		*.63*	*.30*	*.35*	*.40*	--	--
3. Work Life (12 items)	.71 (.18)	.58	.47	.47		*.39*	*.38*	*.47*	--	--
STEM										
4. Anger (12 items)	4.11 (.26)	.45	.46	.30	.47		*.74*	*.79*	*2.51 (.26)*	*.76*
5. Fear (8 items)	4.69 (.23)	.48	.35	.33	.45	.34		*.78*	*2.60 (.35)*	*.79*
6. Sadness (14 items)	4.91 (.31)	.74	.44	.48	.51	.44	.48		*2.62 (.33)*	*.81*

Note. Some items were removed from subsets of the multiple-choice STEM and the STEU to improve reliability (no items were removed from the ratings-based STEM).

Table 3.10 shows the factor loadings and fit indices for one-factor and two-factor models that do or do not differentiate between Understanding and Management. For analyses involving the multiple-choice STEM, χ^2 fit indices indicate no significant improvement in fit when Understanding and Management form two separate factors. In fact, the AIC is slightly better for the one-factor than two-factor model ($\Delta AIC = 1.675$). Both one-factor and two-factor models are over-fitted (i.e., $\chi^2 < df$) and the models are quite similar, since the estimated correlation between the Understanding and

Management factors was almost 1.00 (r = .96). STEM and STEU are measuring the same construct when both are in multiple-choice formats.

Results were quite different for analyses involving the ratings-based STEM. A two-factor model clearly fit the data better than a one-factor model: $\Delta\chi^2$ was significant; ΔAIC = 53.930; and other fit indices were acceptable for a two-factor but not a one-factor model (CFI and GFI > .95; RMSEA < .06; Hu & Bentler, 1999). Factors were still correlated quite highly (r = .57) but not so highly as to indicate they were measuring exactly the same thing. Essentially, STEU and STEM are measuring different constructs when in different response formats but the *same* construct when in the same response format.

Table 3.10

Factor Loadings, Factor Correlations, and Fit Indices for Nested Models Analysis Differentiating between Understanding and Management (Analyses Conducted Separately for Multiple-Choice and Ratings-Based STEM)

	Multiple-Choice STEM (N = 108)		*Ratings-Based STEM (N = 88)*	
	1-factor	*2-factor*	*1-factor*	*2-factor*
Factor Loadings				
STEU (abstract)	.643	.647	.471	.725
STEU (personal life)	.608	.615	.457	.773
STEU (work life)	.739	.748	.524	.808
STEM (anger)	.615	.619	.853	.858
STEM (fear)	.591	.600	.847	.854
STEM (sadness)	.725	.739	.915	.920
Factor correlation		.962		.570
Fit Indices				
χ^2 and *df*	6.660 (9)	6.335 (8)	61.620 (9)	5.690 (8)
RMSEA	.000	.000	.259	.000
CFI	1.000	1.000	.812	1.000
GFI	.981	.981	.792	.979
AIC	30.660	32.335	85.620	31.690
$\Delta\chi^2$ (1 *df*)	0.325		55.930	

4. Discussion

Three of the five hypotheses received full support and the other two were not strongly supported. Specifically, there was full support for individual differences in rating scale use, response format acting as a "method effect", and the STEU and STEM's distinction from personality. However, there is some doubt as to whether the STEU and STEM met the correlational criteria for an intelligence, and whether they related meaningfully to criterion variables. Although the three EI tests were strongly related, correlations between vocabulary and EI test scores were higher than those among EI tests themselves, which is a problem for discriminant validity. STEU and STEM scores related to externally-oriented thinking, but only the STEM incrementally predicted life satisfaction (and only to the retrospective component), and only the STEU incrementally predicted academic achievement. Interactions between EI and vocabulary were found for the STEM in predicting difficulty describing and identifying feelings. No EI measures related to psychological distress. As hypothesised, both STEU and STEM were clearly independent from personality with small relationships to Agreeableness. However, the expected relationship to Openness to Experience was not observed. The paragraphs below provide further discussion of both these issues.

4.1 Reviewing Attempts at Standards-Based Scoring of the STEU and STEM

The use of coping research as a basis for scoring the STEM failed, whereas using Roseman's (1996) appraisal theory to score the STEU was reasonably successful. By looking at the fundamental differences between these two approaches, some useful generalisations about standards-based scoring approaches can be made. Firstly, the structure of emotions used to score the STEU was based on *one* conceptualisation of the important variables (emotions, and aspects of the situation) whereas the coping literature was drawn from many slightly different paradigms for viewing coping mechanisms. Drawing conceptual correspondences between different research paradigms may have introduced a source of error. Secondly, STEU item stems and response options were explicitly based on Roseman's structure of emotions, whereas the coping-based scoring was applied to items already developed according to the SJT paradigm. This may have introduced a further source of error in that

STEM response options did not clearly represent different types of coping mechanisms, nor item stems different types of situations. It seems that the best way to develop standards-based scoring systems is to use a single comprehensive theory (based on empirical evidence) to develop all three of item stems, options, and scoring key.

4.2 Response Formats and Scoring Types

4.2.1 Individual Differences in Rating Scale Use

There were clearly individual differences in the use of the ratings-scale for the STEM, and to a lesser extent for the Stories test. Biases toward using the top or bottom of the scale were related to well-known personality traits, to externally-oriented thinking, to gender (for the STEM) and to cultural group identification (for the Stories). The relationship between high raw ratings on the STEM and Agreeableness may indicate an acquiescent form of responding, where Agreeable participants tend to agree that strategies are effective in general (such that ratings have different meanings depending on levels of Agreeableness). This may be a general problem with ratings-based tests of this type. Given individual differences in rating-scale use, profile scoring would seem to be more valid than non-profile scoring.

4.2.2 Response Format (Multiple-Choice vs Ratings-Based) as a Method Effect

Results supported the suspected "method effect" where Emotional Understanding and Emotion Management related far more strongly when both tests are multiple-choice, then when Understanding tests are multiple-choice and management tests use rating-scale formats. Both zero-order correlations and structural equation models indicated a method effect. Constructs measured by the STEU and STEM were almost identical when both used multiple-choice but clearly separable when the STEM used ratings-based response formats. Although sampling error and small sample sizes may have affected results, sampling error differences seem unlikely to explain such differences. Given greater variability and reliability of tests for the group attempting ratings-based STEM, a greater relationship between the STEM and STEU might be expected for this group (as correlation coefficients would not be attenuated as much by low reliability).

The correlation between STEU and STEM latent factors was .96 when both were multiple-choice, but .53 when STEM was ratings-based. In structural models of the MSCEIT, inter-branch factor correlations between Understanding and Management factors are quite high even though Understanding and Management tests have different response formats, with estimates ranging from .76 (Mayer et al., 2001)[23] to .84 (Day & Carroll, 2004). Given the very high degree of relationship between MSCEIT Understanding and Management factors, it is possible that when method effects are removed (i.e., if both branches used the same scoring format) that Management and Understanding might emerge as the same thing. Empirical evidence that Understanding and Management are separate constructs may well be an epiphenomenon.

4.2.3 Response Formats and Scoring: Implications for Future Research

Results obtained in this study emphasise two important issues that should be considered in further research: (1) profile scoring, and (2) response-format method-related variance. Researchers using items with ratings-based response formats should be aware of the possibility of individual differences in the use of the response format, and check that these are not contributing to the variation in the construct measured. If ratings-based response formats are used (and the current study does imply that these are much more reliable than multiple-choice response formats) then profile-based methods of scoring that account for individual differences in rating-scale use should be applied.

Interpretations of the relationship between constructs must consider that different response formats might result in method-related variance. Ideally, indicators of the same construct could use several possible response formats, allowing a design after Campbell and Fiske's (1959) multi-trait multi-method assessment of reliability and validity. However, given the practical difficulty of such a design (the vast majority of existing cognitive instruments do not employ a ratings-based format),

---

[23] Inter-branch correlations were not reported in the paper, but this value of $r = .76$ was deduced from the correlation matrix and model specifications given. By specifying that the correlation between branches 1 and 2 be equal to that between branches 3 and 4 Mayer et al. (2001) probably underestimated the correlation between branches 3 and 4, as zero order correlations between tests were considerably higher for 3 and 4 than for 1 and 2.

method variance might be controlled for by making sure that all tests are of the *same* response format. This possible response-related method effect was controlled for in such a way in Chapter 5, when examining the distinctiveness of EI from intelligence.

4.3 Support for the Correlational Criteria for an Intelligence

There was partial support for STEU and STEM meeting correlational criteria. High correlations between STEU and STEM, and lower correlations between these and Stories remained when vocabulary was controlled (i.e., the relationship between them was not entirely explained by verbal ability). However, both the STEU and multiple-choice STEM related more strongly to vocabulary than to Stories (even when the low reliability of the Stories test was controlled). This may be a problem with discriminant validity given that EI should be distinct from other forms of intelligence.

High correlations with vocabulary may relate to either: (1) the verbal ability required to comprehend items; or (2) to a genuine relationship between Strategic EI and Gc. In the first case, managing and understanding emotions may constitute a new type of intelligence where test scores relate strongly to vocabulary only because of method variance (i.e., items require a high base level of reading skill). In the second case, managing and understanding emotions may constitute a form or subset of acculturated knowledge (as Matthews et al., 2002 and Roberts et al., 2000 suggest). Further research reported in Chapters 5 and 6 involved a fuller empirical array of crystallized intelligence markers to determine which whether the relationship was with Gc generally or a narrower reading skill particularly.

4.4 Relationships of EI to Criterion Variables

4.4.1 Relationships of EI to Alexithymia, Life Satisfaction and Emotional Distress

The hypothesis that STEU and STEM scores would relate to alexithymia, life satisfaction, and psychological distress also received only partial support. Both STEU and STEM related to the externally-oriented-thinking style facet of alexithymia even after personality and intelligence were controlled. However, there were no significant relationships with other aspects of alexithymia

(difficulty identifying or describing feelings). The multiple-choice STEM but not the STEU or the ratings-based STEM related to retrospective but not current life satisfaction. A possible reason for the retrospective-only relationship may be the use of positive re-appraisal as an effective emotion management strategy. Given that those with high levels of emotion management use more positive re-appraisal (Lyons & Schneider, 2005), the relationship between retrospective life satisfaction and emotion management may be due to positive re-appraisal of past events. Certainly, such an outcome supports the delineation of life satisfaction into current and retrospective areas as a useful one. The STEU did not relate to life satisfaction, and neither STEM nor STEU related to the general health questionnaire.

The TAS-20 scores are self-ratings, whereas the STEU and STEM are performance-based, which may account for some of the lack of correlation between these measures. Dawda and Hart (2000) found that trait EI (EQ-i) scores were much more strongly related to TAS-20 scores than to more objective other-report measures of alexithymia, and the TAS-20 has subsequently been used as a proxy measure of trait EI (Barchard & Hakstian, 2004; Palmer, Donaldson, & Stough, 2002). Since trait and performance-based EI have a small to zero relationship (e.g., Brackett & Mayer, 2003), it may have been more appropriate to use alternative measures to assess alexithymia rather than trait EI (e.g., the older Beth Israel Questionnaire [BEQ]; Sifneos, 1973).

Both the STEU and STEM were clearly unrelated to psychological distress as measured by the GHQ. However, the extreme scores on the GHQ (where participants were well above the comparison average, and nearly half the sample above the clinical cut-offs) indicate that this measure may not have adequately measured participant's state psychological distress. Administering this symptom-checklist-like measure directly after the OCEANIC and TAS-20 rating scales may have introduced demand characteristics such that participants did not check the "nothing wrong" option 20 times, even if nothing was wrong. The following chapter (Chapter 4) re-examines the relationship of the STEU and STEM to states of psychological distress using different instruments and (importantly) not administering the checklist following personality measures, so as to avoid this potential problem.

4.4.2 Relationships of EI to Academic Achievement

Only the STEU (and not the STEM) incrementally predicted academic performance, and only for psychology grade (not for overall average grade). Stronger prediction of academic achievement from Understanding than Management is consistent with prior research (Barchard, 2003). It is possible that understanding social and emotional dynamics (i.e., the emotions occurring in a situation) might relate to content knowledge in a psychology course (so that the EI relationship is to course content rather than adapting to the emotional demands of university life). Some existing research has found SJTs assessing interpersonal skills predict grades in courses with interpersonal components but not other curricula, supporting this idea (Lievens, Buyse, & Sackett, 2005). However, personality and vocabulary were also stronger predictors of psychology grade than overall grade implying that the difference is more likely due to the greater consistency in psychology grade as a variable. Since overall grade is composed of different subjects for different people (which may have different demands and levels of difficulty), there may be more "error" in overall grade than psychology grade. Laird (2005) points out that there is ample evidence that grading practices may be considerably different in humanities courses compared to engineering and science courses, supporting this idea.

The incremental validity of the STEU in predicting an additional 8% of the variation in grades over personality and vocabulary appears impressive, but may be an over-estimate of the true state of affairs. Only vocabulary was controlled in the regression, such that any overlap between the STEU and fluid intelligence (or other aspects of crystallized intelligence) would show up as STEU-related prediction. Chapter 6 examines the STEU and STEM's prediction of academic achievement after controlling for personality, and both fluid and crystallized intelligence. However, as a first approximation, prediction of academic achievement from the STEU does provide some important evidence of predictive validity.

4.5 Relationships between EI and Personality

As hypothesised, STEU and STEM scores were clearly independent from personality, showing no strong relationships with any dimension of the five factor model. However, small

relationships with Agreeableness were observed, consistent with results for the MSCEIT (Brackett & Mayer, 2003; Brackett et al. 2004; Lopes et al., 2004; Lopes et al., 2003; Schulte et al. 2004). Given that EI is a form of intelligence, and intelligence relates to Openness, the STEU and STEM's lack of relationship to Openness may appear difficult to interpret (see Ashton et al., 2000; Bates & Shieles, 2003; Harris, 2004 for evidence of an Openness-Gc relationship). Prior research on the MSCEIT has also mostly found support for an EI-Openness relationship (see Brackett & Mayer, 2003; Brackett et al., 2004; although Lopes et al. [2003] provide an exception where there was a significant negative correlation with Openness). However, the brief measure of Openness used in this study was mostly representative of "Openness to Ideas", rather than Openness' other facets (Fantasy, Aesthetics, Feelings, Actions, and Values), and EI may relate to "Openness to Feelings" more strongly than to other facets. As with the vocabulary-EI relationship, more detailed measures of personality which allow examination of relationships at the facet rather than broad dimension level would be useful in resolving this issue. This research was subsequently planned and is presented in Chapter 5.

4.6 Limitations of this Study

A primary limitation of this study was that the sample sizes were quite low for structural equation modelling ($N = 113$ and $N = 94$ respectively). Replication with larger samples is thus needed to confirm the current findings. In Chapter 4, the confirmatory factor analysis involving multiple-choice measures was repeated with a larger sample to address this issue. A second limitation of the current study was the non-random allocation of subjects to group A and group B, which may have confounded results (i.e., it may be a rural versus city campus student distinction that lead Management and Understanding to differentiate, rather than a multiple-choice versus ratings based response format distinction). As such, findings that Management and Understanding differentiate when measured in different response formats (but not the same response format) are in need of replication.

4.7 Concluding Comments for Chapter 3

In general, the STEU and STEM's construct validity as measures of performance-based EI seems reasonable on most counts. However, the predictive validity of both tests was questionable, requiring further support. Both STEU and STEM scores related to each other, to Stories, to the externally-oriented-thinking aspect of alexithymia and were relatively independent of personality. The STEU also incrementally predicted psychology grades and the STEM incrementally predicted retrospective life satisfaction. Relationships with vocabulary may be too high for discriminant validity, and lack of relationships with two aspects of alexithymia might be problematic, but on the whole the tests were performing reasonably. With some provisos about the overlap with existing forms of intelligence, the STEU and STEM do appear to be relatively valid measures of EI, and will be used as such in further studies (subject, of course, to further evidence to the contrary). The possibility of response format acting as a source of variance has important implications for the structure of EI, which so far has been determined from the MEIS and MSCEIT tests where Understanding and Management measures have different response formats. The following chapter re-examines the distinction between Management and Understanding when both are in the same response format, using an adult rather than a student sample. The next chapter also re-examines whether the STEU and STEM can predict states of psychological distress, using a different measure to the GHQ and not administering the measure following personality tests (which may have caused problems in the current study).

Chapter 4

Does EI Predict Psychological Distress? Concurrent Validity and Reliability of the Situational Test of Emotional Understanding (STEU) Situational Test of Emotion Management (STEM) and in a Non-Student Sample

1. Introduction to Chapter 4

The previous two chapters outlined the development and validation of the STEU and STEM in a sample of psychology undergraduate students. The current study's purpose is to provide further validation of the STEM and STEU, examining four key issues: (1) the STEM and STEU's reliability in a non-university sample; (2) the differentiation of Understanding from Management when both are given in multiple-choice format; (3) the relationships of STEM and STEU scores to state anxiety, depression, and stress; and (4) their prediction of life-style variables indicative of success and emotional health. These issues are described in more detail below.

1.1 The Importance of a Non-University Sample

In items assessing reasoning ability, standard practice is to use stimuli that have an equal degree of familiarity to all participants, ensuring that sources of variation due to irrelevant demographic or life experience variables are minimised (e.g., Jensen, 1980). However, this is difficult, if not impossible, to control for in situational tests and may actually interfere with measuring the construct of interest (i.e., if the construct of interest includes knowledge about particular types of situations, there may well be group differences caused by different levels of experience in these types of situations). STEM and STEU items are based on hypothetical situations occurring in work-place and personal life contexts, and test-takers' different levels of experience in these different contexts may affect their test scores.

Although the STEM and STEU were reasonably reliable for undergraduate psychology students with relatively homogeneous life experience, this might not be the case in other groups who differ in life experience, and are familiar with different types of situations. Landy (in press) suggests that the almost exclusively student samples used in social intelligence research may strongly affect the research outcomes. He uses this as a cautionary tale for EI, pointing out that social and emotional abilities (particularly understanding of social or emotional situations) may change drastically at different ages. In addition, the exclusive use of psychology undergraduates as research samples may be problematic, given that people who choose to study psychology may well differ from the general population in terms of their interest and attention to emotional and inter-personal phenomena. As such, the reliability of the STEU and STEM in undergraduates found in Chapter 3 may not be an adequate indication that these tests are reliable in other groups, which is why this study also assesses the reliability of these measures in a non-university sample.

1.2 Relationship of Strategic EI to States of Distress

Another measure of life success outside the areas cognitive ability tests typically predict is effective coping with stressors. If people can understand emotional situations, and know the best strategies for managing them, then they ought to cope better via the use of more appropriate strategies (Salovey et al., 2000). Taylor (2001) suggests that stress is avoided by high EI individuals who can balance or contain their emotional responses and thus protect themselves from the most adverse effects of stressful situations. Matthews et al. (2002) outline the link between higher EI and better coping in more detail, suggesting six possible factors mediating the EI-coping relationship. According to Matthews et al., higher EI people cope better by: (1) avoiding stressful encounters; (2) having richer coping resources; (3) perceiving and appraising situations more constructively; (4) adaptively repairing and regulating emotions (and being more self-efficacious in this regulation); (5) having greater emotional skills (e.g., the disclosure of trauma) and (6) having greater competence and flexibility in their use of effective coping strategies. Points 4, 5, and 6 are conceptually related to the

Management construct assessed by the STEM, and point 3 (the correct appraisal of emotional situations) to the Understanding construct assessed by the STEU.

An obvious result of better coping with stress is lower state distress. Although Chapter 3 found no relationship between psychological distress (as measured by the GHQ-20) and either STEU or STEM scores, the particularly high average score on the GHQ indicated that the instrument may not have accurately measured distress in this instance. This chapter uses an alternative measure: the DASS (Depression Anxiety Stress Scales, Lovibond & Lovibond, 1995), which differentiates between states of depression, anxiety, and stress, rather than looking at undifferentiated distress. Some existing research using the MEIS and MSCEIT (outlined below) has found that performance-based EI does predict lower states of depression, anxiety, and stress, indicating that it is worth re-examining the relationship between the distress and the STEU and STEM.

Matthews, Emo, Funke, Zeidner, Roberts, Costa, and Schulze (submitted) report that MSCEIT scores relate to lower states of distress and worry and less avoidance-based coping, even after controlling for Neuroticism. However, after participants were exposed to a stressful task, MSCEIT scores no longer predicted distress, although they still predicted worry. Bastian et al. (2005) found that the MSCEIT did *not* incrementally predict coping (as measured by the COPE; Carver, Scheier, & Weintraub, 1989) after personality and intelligence were accounted for, but *did* incrementally predict lower state anxiety. Ciarrochi et al. (2002) found no relationship between the Stories task from the MEIS and state depression. Based on these results and Matthews et al.'s (2002) suggestions, it is hypothesised that the STEU and should predict state anxiety more strongly than stress or depression, although relationships to all three indicators of distress might be expected. These relationships are also expected to be stronger for the STEM, as the construct is more strongly related to Matthews et al's (2002) potential mediators of stress.

1.3 Predicting "Real Life" Measures of Success

Existing research on EI has often focused on life-style variables as validity indicators of EI. Brackett and Mayer (2003) found that the MSCEIT predicted social deviance (e.g., fighting) but not

tobacco, alcohol, or drug use in an undergraduate sample, although Brackett et al. (2004) found that Understanding and Management scores predicted illegal drug use for males (also in an undergraduate sample). However, scores obtained from the MSCEIT's predecessor (the MEIS) did predict tobacco and alcohol use and predictive factors for smoking risk in teenagers (Trinidad & Johnson, 2002; Trinidad et al., 2004). MEIS and MSCEIT scores also predicted relationship quality, fewer negative interactions with close friends, positive relations with others, and social support from parents (Ciarrochi et al., 2000; Lopes et al., 2003).

Most of these studies assessed these life-style variables using ratings of attitudes, sentiments, or self-descriptors. An exception to the use of ratings scales is Brackett et al. (2004), whose life-space approach instead asks for verifiable facts. This type of information is thought to be less susceptible to faking or response bias, and thus might be considered more valid (Shaffer, Saunders, & Owens, 1986). Brackett et al. found that Strategic EI scores (i.e., Understanding and Management) predicted both positive and (lack of) negative relationships with friends. Based on such reasoning, the current study collected some basic verifiable facts indicative of positive or healthy life-styles: the frequency of smoking and drinking in a typical week, the number of visits to family and friends in a typical month, and the membership in a community group, job title, and income. Although such measures are fairly indirect indicators of life success they were chosen for their convenience, being quick and easy to collect.

Strategic EI (and particularly Management) should relate to lower levels of smoking and problematic drinking through the use of more functional rather than dysfunctional coping strategies. For example, Zeidner, Matthews, and Roberts (in press) report that MSCEIT scores predicted more use of active coping strategies and less use of dysfunctional strategies such as alcohol use. EI should also relate to greater contact with family and friends through the ability to maintain and create social support networks (Salovey et al., 2000; Zeidner et al., in press). The ability to understand the emotional content of inter-personal situations (Understanding) as well as knowing the most effective responses in such situations (Management) would be useful in a large variety of jobs in terms of

gaining help, clients, promotions, or other resources. Although income and job status are fairly indirect measures of success in the workplace it is expected that STEU and STEM scores may relate to these, as well as to drinking and smoking behaviour and contact with family and friends.

1.4 Differentiating Emotional Understanding from Emotion Management

As Chapter 3 demonstrated, the existing factor-analytic evidence differentiating Understanding from Management may be based at least partially on method variance. The difference may be between ratings-based and multiple-choice response formats rather than between Understanding and Management constructs. Replication of these results will be attempted in the current study, where both STEU and STEM have been given in multiple-choice format. A structural examination of whether STEU and STEM item parcels load onto the same factor or onto two different factors representing Understanding and Management will be undertaken. This issue has important implications for the four-branch hierarchical model of EI. The four-branch model is currently based on empirical evidence derived solely from the MEIS and MSCEIT (where Understanding and Management are in different formats), and therefore may be contaminated by method variance.

1.5 Summary of Hypotheses

Hypothesis 1. STEU and STEM scores will be still be reliable and correlate strongly with each other in a non-undergraduate sample.

Hypothesis 2. STEU and STEM scores will relate to states of anxiety, depression, and stress, with the strongest relationships between anxiety and the STEM.

Hypothesis 3. Structural results from Chapter 3 will be replicated, with STEU and STEM item parcels better explained by *one* EI factor than by two separate factors of Understanding and Management.

Hypothesis 4. STEU and STEM scores will relate to “success” related lifestyle variables, predicting income, job status, smoking, and drinking behaviour, frequency of visits to friends and family, and membership in community organisations.

2. Method

2.1 Participants

Participants were volunteers recruited from the inner city, inner west, and southern regions of Sydney. Advertisements were placed in local papers, a magazine freely distributed in Sydney's CBD, and fliers and circulars were placed in shops, libraries, and community notice-boards. As an incentive to participate, participants were offered feedback on their scores (an example of such feedback in given in Appendix 4.1). As an ethical requirement from Sydney University's Human Research Ethics Committee, participants whose scores were 2 standard deviations above the population mean on DASS anxiety or stress did not complete the STEU or STEM (as the possibility of receiving negative feedback might be upsetting for vulnerable participants). 14 participants were excluded under these conditions: The computer program did not allow them to access the STEU or STEM, but instead sent them to a page informing them that the study could not use them at this time.

The sample was unlikely to be representative of the Australian public, as participation required computer and internet literacy, and was drawn mainly from central Sydney rather than the suburbs or regional areas. The purpose of using community volunteers was not to obtain a representative sample but a sample that differed from the undergraduates in their life experience.

149 participants (107 female) took part in this study.[24] Participants' ages ranged from 18 to 59 with a median of 33 (M = 35.33, SD = 11.03). Of the 138 people who reported their weekly income, the mean was \$1124.29/week (SD = \$774.03, median = \$1000.00, maximum = \$5000, minimum = \$68), significantly more than the national average of \$766.10 (as of November, 2004; Australian Bureau of Statistics, 2005, $t_{137} = 5.284$, $p < .001$). Even when the five highest income earners were excluded (as their standardised scores for income were greater than 3.00), this sample still earned significantly above the national average (M = \$1019.40, SD = \$548.04, $t_{132} = 5.120$, $p < .001$). 112

[24] This figure excludes the 14 people who scored more than 2 SDs above the reported mean on anxiety or depression. All biodata statistics exclude these 14, although they were included in the calculation of reliability of the DASS sub-scales.

people worked full-time (105 of these reported their income, M = \$1299.62, SD = \$790.72, compared to a national average of \$1021.90; $t_{104} = 3.599$, $p < .001$), 25 part-time (of whom 23 reported their income, M = \$599.70, SD = 287.31).[25] Even given that selectively missing data might decrease the mean income, the current sample does appear to be earning well above the Australian national average. In addition, when job titles were matched to the ANU social status scales (which range from 0 to 100; see McMillan & Jones, 2000), this sample's job status mean of 56.72 (SD = 15.26) was significantly different from a middle value of 50 ($t_{138} = 5.187$, $p < .001$).

110 participants identified as primarily Australian or Anglo-Celtic (104 as *only* Australian or Anglo-Celtic, with no identification to any other cultural groups). Of the 35 participants identifying with other cultures, 19 identified with Asian cultures, and 16 with European (non-UK) cultures. 24 were non-native English speakers, who had been speaking English for between 3 and 46 years (M = 22.79, SD = 11.65).

The majority of participants (136) completed high school, with all but 6 completing a year or more of post-secondary education (mean years of post-secondary education = 4.62, SD = 2.63). 58% of Australians aged 25 to 64 have a post-secondary qualification (Australian Bureau of Statistics, 2005). Making the conservative estimation that four years of post-secondary education equals a qualification, 68% of the current sample have post-secondary qualifications, significantly more than in the general population ($\chi^2(1) = 5.561$, $p = .018$). 67 participants had never married, and 67 were in marriage-like relationships (26 in de-facto relationships and 41 legally married), 10 were divorced, and 4 were separated but not yet divorced.[26] 84 participants reported belonging to a community-based hobby or sports group.

---

[25] Based on the weekly incomes and job titles, it appeared that 9 participants reported their annual income, rather than weekly income. These 9 income reports were adjusted by dividing by 52 before analysis was undertaken.

[26] 1 person did not report marital status, 2 did not report high school level completed, and 1 did not report years of post-secondary education.

2.2 Test Battery

2.2.1 Situational Test of Emotional Understanding (STEU)

The full 42-item version of the STEU was administered to participants. The development of the STEU is described in Chapter 2 of this thesis, and its validation in Chapter 3. Appendix 2.1 reproduces all instructions and items of the STEU. Cronbach alpha reliability was .71 in the undergraduate sample in Chapter 3.[27] The STEU is scored according to standards derived from Roseman's (2001) theory of the structure of emotions.

2.2.2 Situational Test of Emotion Management (STEM; Multiple-Choice Format)

The STEM was given in multiple-choice format and scored according to expert proportions. In an attempt to improve the reliability of the STEM, the 4 least reliable items were removed from the scale, and 13 additional items with low reliability were modified. Appendix 4.2 describes the item analysis used to select these unreliable items in more detail. The 13 modified items were re-scored by three of the experts from the EI research group (there was complete agreement on the best answer for 9 of these items, 2/3 agreement for 3 items, and disagreement on 1 item). The total test was thus 40 items long, 13 "new" items and 27 "old" items.

2.2.3 Depression Anxiety Stress Scale (DASS, Lovibond & Lovibond, 1995)

The 21-item version of the Depression Anxiety Stress Scale (DASS) consists of 21 items rated on a 4-point scale from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time). There are 7 depression items (e.g., "I felt downhearted and blue"), 7 anxiety items (e.g., "I felt scared without any good reason") and 7 stress items (e.g., "I found it difficult to relax"). Antony, Bieling, Cox, Enns, and Swinson (1998) reported DASS-21 Cronbach alphas of .94 for Depression, .87 for Anxiety, and .91 for Stress.

---

[27] Due to a technical problem, participants' answers to question 35 (distress in a personal life context) were not recorded, so that the STEU was in effect 41 items long rather than 42.

2.2.4 Demographic and Lifestyle Questions

There were 17 questions in the survey assessing demographics and life-style factors. Participants reported their age, sex, marital status, ethnic/cultural group(s) they identified with, job title (which was later matched to status scales), weekly income, hours of work (full-time or part time), number of times per month they visited friends, number of times per month they visited family, and whether they were a member of a community group. Participants also reported their tobacco use as one of five categories: never smoked, no longer smoke, smoke less than once a week, smoke more than once a week but not daily, and smoke daily. Similarly, they reported their alcohol use as one of four categories: never drink alcohol, drink alcohol less than once a week, drink alcohol more than once a week but not daily, and drink alcohol daily.

2.3 Procedure

After reading the initial advertisement or circular, participants emailed the author stating their willingness to participate and the author responded with a form email giving the URL and instructions for participating. Participants then read through a page of instructions and completed demographic and lifestyle questions, followed by the DASS, STEU, and multiple-choice version of the STEM. In line with confidentiality concerns, participant's scores were identified only by a code number. The only unique identifying information collected was their email addresses (for the purposes of feedback), which was kept separate from the main data file. Participants were emailed feedback outlining what the tests measured and whether their performance was "typical" (within the 0-70th percentile), "superior" (70th to 80th percentile), "very superior" (80th to 90th percentile) or "excellent" (90th to 99th percentile) compared to the sample from Chapter 3. An example of such feedback is shown in Appendix 4.1. This study was approved by the Sydney University Human Research Ethics Committee.

2.4 Analyses

There were four stages of analysis in this study. After assessing the internal consistency of the STEU and STEM in the current sample, structural analysis of the STEU, STEM, and DASS was undertaken, and factor scores compared to life-style and demographic variables.

Analysis 1: Calculation of Reliability. Cronbach alpha reliability will be calculated for the STEU and STEM using the current sample, to compare these reliability estimates with those obtained in Chapter 3 (a psychology student sample).

Analysis 2: Structural Examination of the STEU and STEM. The distinction between Understanding and Management will be tested via structural equation modelling of STEU and STEM subsets. A one-factor model (which all STEU and STEM subsets load on) will be compared to a two-factor model (where STEU subsets load on an Understanding factor and STEM subsets load on a Management factor).

Analysis 3: Structural Examination of the DASS. The three-factor structure of the DASS will be tested with confirmatory factor analysis.

Analysis 4: Structural Examination of the DASS, STEU, and STEM. Analyses 2 and 3 will be combined such that DASS and Strategic EI factors are included in the same model, allowing a test of the relationship between Strategic EI and state depression, anxiety and stress.

Analysis 5: Relationships between Strategic EI and Demographic/Lifestyle Variables. Analyses 2 and 3 will be combined such that DASS and Strategic EI factors are included in the same model, allowing a test of the relationship between Strategic EI and depression, anxiety and stress.

3. Results

3.1 Reliability and Descriptive Statistics

Table 4.1 shows the reliabilities and descriptive statistics for the STEU, STEM, DASS, and continuous life-style variables. There were no significant mean differences between men and women on any of these scores. Reliabilities are acceptable for the DASS scales and STEM, but lower for the STEU. Removing items did not help the STEU's reliability. For the STEM, the 13 new items are not

adding substantially to the reliability, as the 40-item version is only slightly more reliable than the longer version. This sample's STEU and STEM scores are much higher than those obtained by the psychology undergraduate sample in Chapter 3, with a medium to large effect size for the STEU and small but non-trivial effect size for the STEM. Given this difference, it seems likely that lower reliabilities obtained here than in Chapter 3 are at least partly due to restriction of range. However, lower reliabilities may also indicate that the STEU and STEM are not transferable to a non-university student population. A further implication of this sample's higher mean scores on the STEU and STEM is that there may be age trends in EI similar to those found for Gc (e.g., Horn & Cattell, 1967). DASS scores were reasonably internally consistent, with Cronbach alphas similar to but slightly lower than those reported in Lovibond and Lovibond (1995). Mean differences from the comparison sample were all less than .20 (Cohen's [1988] cut-off for "small").

Table 4.1

Descriptive Statistics and Reliabilities for DASS, STEU, and STEM (Comparison Parameters from Chapter 3 and the Significance of the Difference from these Shown on Right)

	Current Sample			*Comparison Sample*			*d*
	α	*Mean*	*SD*	*α*	*Mean*	*SD*	
Strategic EI [a]							
1. STEU	.51	.70	.11	.72	.61	.13	.74
2. STEM (40 items)	.64	.65	.08				
2b. STEM ("old" 27 items)	.59	.60	.08	.71	.58	.10	.23
DASS-21 [b]							
4. Depression	.91	3.63 (2.66)	4.21 (2.67)	.91	3.60	3.27	.01
5. Anxiety	.73	2.69 (2.30)	2.89 (2.25)	.81	2.62	2.42	.03
6. Stress	.83	5.93 (5.38)	3.83 (3.29)	.89	5.27	3.47	.19
Lifestyle Variables							
7. Age	-	35.33	11.03				
8. Annual income	-	1124.29	774.03				
9. Job status	-	56.72	15.26				
10. Years of education	-	16.48	2.79				
11. No. monthly visits to family	-	3.02	3.37				
12. No. monthly visits to friends	-	6.89	6.14				

Note. DASS parameters in brackets are for the sample excluding the 14 participants scoring > 2 SDs above the mean on anxiety or depression; *d*-scores are calculated using the whole sample, including these 14.

[a]Comparison sample for the STEU and STEM from Chapter 3 of this thesis. [b]Comparison sample for the DASS-21 from Lovibond and Lovibond (1995). Since Lovibond and Lovibond report parameters for the 42-item DASS, these values are halved for comparison purposes, as instructed in their paper.

3.2 Structural Model of EI

The distinction between Understanding and Management was tested with nested models in a similar design to Chapter 2. However, this chapter used item parcels rather than subsets. A one-factor model where all indicators loaded on a single Strategic EI factor was compared to a two-factor model where STEU and STEM item parcels loaded on Understanding and Management factors respectively.

14 item parcels for the STEU were created, each consisting of three items dealing with the same emotion. For example, items representing relief in a work-context, relief in a personal life context and relief in an abstract context were combined to form one item parcel of "Relief". STEM items were combined by the general theme they expressed. Themes included "Goal Striving Impeded" (for which there were 3 items); "Unfairness" (6 items); "Rudeness" (6 items); "Anticipation" (5 items); "Role Change" (3 items); "Health Concerns" (3 items); "Workplace Isolation/Loneliness" (4 items); "Personal Life Loneliness" (5 items); "Parting" (3 items); and "Loss" (2 items; see Table 2.2 in Chapter 2, Part II, Section 1.3 for a full description of these general themes). Correlations between the indicators are shown in Appendix 4.3. Indicators that did not significantly load on factors in either the one-factor or two-factor model were excluded from the analysis (6 STEU and 2 STEM parcels were excluded, resulting in 8 indicators for each latent variable). This final structure of the two nested models is shown in Figure 4.1.

Figure 4.1

Nested Models Design Differentiating Emotional Understanding and Emotion Management

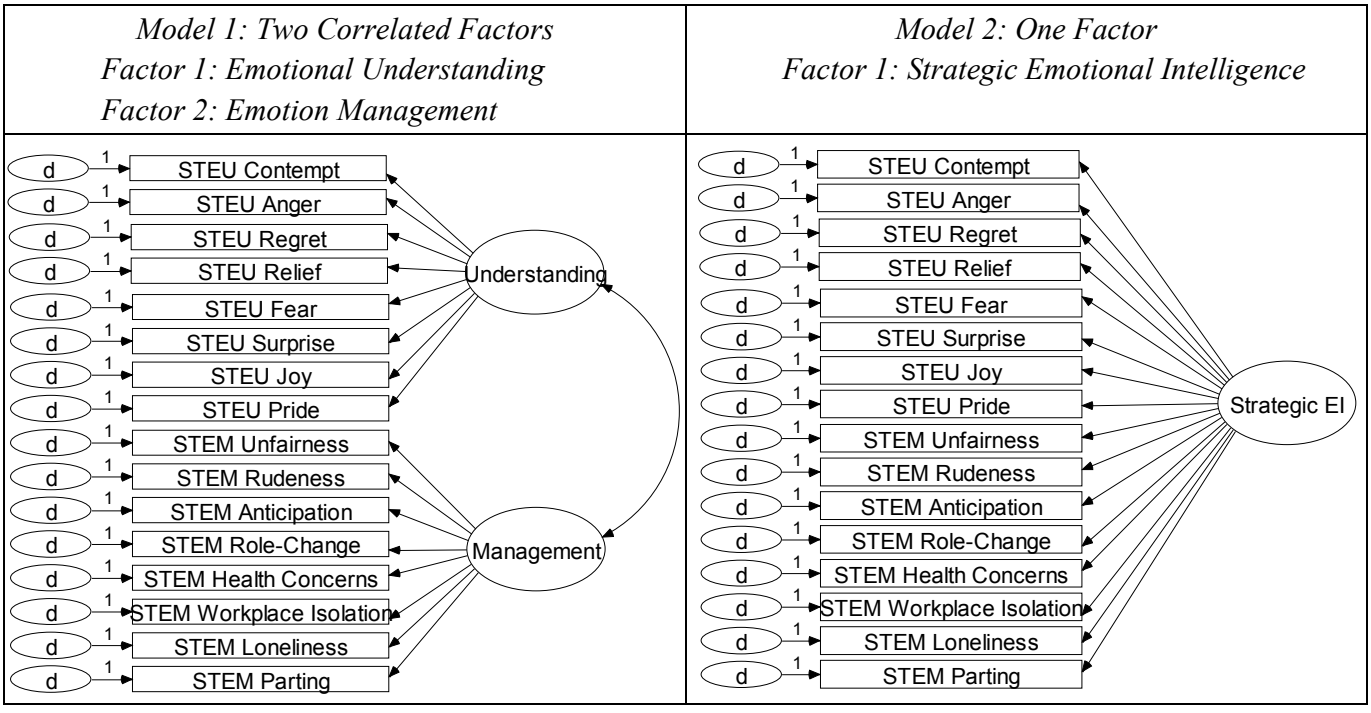


Table 4.2 shows descriptive statistics for the item parcels used and results from the structural equation modelling (factors loadings, factor correlations, and fit indices for both the one and two-factor models). In contrast to Chapter 3, the two-factor model fit significantly better than the one-factor model indicating that Understanding and Management are separate factors ($\Delta\chi^2(1) = 4.93$, $p = .026$; $\Delta AIC = 2.93$). However, the correlation between factors was still relatively high ($r = .722$), such that without a good theoretical reason to differentiate between Understanding and Management, they could well be interpreted as the same thing. By convention, RMSEA values less than 0.05 are regarded as a good fit, and less than 0.08 as an acceptable fit, and values of greater than .90 on the CFI or GFI as good fit (see e.g., McDonald & Ho, 2002). However, Hu and Bentler (1999) discuss the need to change these heuristics to .06 (for the RMSEA) and .95 (for the CFI). The two-factor model showed relatively good fit with RMSEA < .05 and GFI and CFI > .90 (but not > .95). The RMSEA was also good and GFI acceptable for the one-factor model, although the CFI < .90. Results do not support the hypothesis that the STEU and the multiple-choice STEM are measuring exactly the same construct, in contrast to the previous chapter. However, differentiation of Strategic EI into Management and Understanding components does not receive strong empirical support, but rests on a matter of interpretation. The two-factor solution will be used in further analyses for two reasons: (1) it fit significantly better, and (2) examining the relationship between EI and criterion variables separately for Understanding and Management provides additional information on whether these show the same correlates and are therefore likely to be the same construct.

Table 4.2

Nested Models Analysis of STEU and STEM Item Parcels: Descriptive Statistics of Indicators, Factor Loadings, Factor Correlations and Selected Fit Indices.

			1-factor (104 df)	*2-factor (103 df)*	
	Mean	*SD*	*Strategic EI*	*Understanding*	*Management*
STEU Contempt	1.38	.99	.229	.301	
STEU Anger	1.50	.91	.205	.288	
STEU Regret	2.40	.77	.220	.278	
STEU Relief	2.60	.65	.332	.316	
STEU Fear	1.34	.86	.256	.323	
STEU Surprise	2.64	.55	.210	.266	
STEU Joy	2.36	.77	.398	.451	
STEU Pride	2.43	.80	.244	.275	
STEM Unfairness	4.12	.68	.319		.328
STEM Rudeness	3.98	1.04	.618		.620
STEM Anticipation	2.37	.62	.279		.274
STEM Role-Change	1.80	.54	.340		.352
STEM Health Concerns	2.43	.65	.250		.269
STEM Workplace Isolation	2.44	.68	.429		.460
STEM Loneliness	3.30	.79	.361		.375
STEM Parting	2.54	.44	.460		.468
Factor Correlation				.722	
χ^2			120.081		115.051
CFI			.873		.904
GFI			.913		.916
RMSEA			.032		.028
AIC			184.081		181.151

Note. All paths significant at $p < .05$.

3.3 Structural Model of the DASS

A confirmatory factor analysis of the 21-item DASS was undertaken, where the 7 depression items loaded on a depression factor, 7 anxiety items on an anxiety factor, and 7 stress items on a stress factor. This analysis is shown in Table 4.3. Because fit indices were relatively poor (RMSEA > .06; both GFI and CFI < .90), an alternative structure was also tested. Exploratory analyses of all 21 items as well as the 7 anxiety items alone suggested that anxiety was assessing two factors: (1) Physiological Arousal, composed of 4 items from Lovibond and Lovibond's (1995) "autonomic arousal" and "skeletal musculature effects" factors; and (2) the subjective experience of anxiety (Subjective Anxiety), composed of three items from Lovibond and Lovibond's "situational anxiety" and "subjective experience of anxious effect" factors. The isolation of a 4-item physiological arousal factor replicates Duffy, Cunningham and Moore (2005), providing further support that this is a reasonable model.

A second analysis separating anxiety into these two factors (Physiological Arousal and Subjective Anxiety) significantly improved fit ($\Delta\chi^2(3) = 15.132$, $p = .0017$), although indices were still below cut-offs for good fit (i.e., neither of the models provides a very good fit to the data). Both Depression and Stress factors related more strongly to Subjective Anxiety than to Physiological Arousal, and the two aspects of anxiety related only as strongly to each other as they did to Depression or Stress (i.e., there is no evidence that anxiety forms a uni-dimensional construct in the DASS-21). This analysis is also shown in Table 4.3. The four-factor solution will be used in further analyses as this fit better and may also provide more specific information about which aspects of anxiety relate most strongly to the STEU and STEM.

Table 4.3

Structural Models of the DASS-21: Factor Loadings, Factor Correlations and Fit Indices

	3-factor Model (186 df)			*4-factor Model (183 df)*			
	D	*A*	*S*	*D*	*PA*	*SA*	*S*
Factor Loadings							
1. Downhearted/blue	.663			.660			
2. Nothing to look forward to	.695			.695			
3. Life meaningless	.654			.652			
4. Worthless as a person	.627			.627			
5. Unenthusiastic	.635			.637			
6. No positive feelings	.658			.662			
7. No initiative to do things	.508			.507			
8. Heart racing		.335			.515		
9. Dry mouth		.335			.473		
10. Breathing difficulty		.534			.593		
11. Trembling		.304			.323		
12. Panic/make fool of self		.515				.617	
13. Close to panic		.634				.646	
14. Scared without reason		.359				.411	
15. Hard to wind down			.738				.737
16. Difficult to relax			.698				.697
17. Nervous energy			.576				.576
18. Agitated			.604				.606
19. Over-react to situations			.448				.448
20. Touchy			.554				.554
21. Intolerant of obstacles			.523				.523
Factor Correlations							
D		.548	.625		.387	.524	.625
PA (A)			.589			.531	.493
SA							.543
Fit Indices							
χ^2	347.834			332.702			
CFI	.802			.817			
GFI	.825			.833			
RMSEA	.077			.074			
AIC	437.834			428.702			

Note. D = Depression; S = Stress; A = Anxiety; PA = Physiological Arousal; SA = Subjective Anxiety. Items are ordered after Lovibond and Lovibond (1995), p. 339. All paths significant at $p < .01$

3.4 Relationship between Strategic EI and the DASS

The models outlined in Tables 4.2 and 4.3 were combined to test the relationship between Strategic EI and states of depression, anxiety, and stress. Factor loadings for this model are shown in Table 4.4. Fit indices were contradictory: The RMSEA was good (.049); but the GFI and CFI were both poor (.787 and .798 respectively; $\chi^2(614) = 827.877$). However, all paths in the model were significant at $p < .05$ and this model was retained.

Table 4.5 shows the correlation between factors in this analysis. As in the analysis outlined in Table 4.2, Understanding and Management factors are strongly related. However, they relate differently to Depression, Anxiety, and Stress factors. The hypothesis that EI would predict anxiety more strongly than depression or stress was confirmed for Understanding only. Understanding related significantly more strongly to both Physiological Arousal and Subjective Anxiety than to either Depression or Stress (for Physiological Arousal, $z = 2.39$, $p = .017$ for Depression, and $z = 2.04$, $p = .041$ for Stress; for Subjective Anxiety, $z = 4.13$, $p < .001$ for Depression and $z = 3.78$, $p < .001$ for Stress). Management predicted Depression, both anxiety factors, and Stress equally strongly, with no significant differences in the strength of correlations. Generally, relationships between Strategic EI and the DASS are consistent with hypotheses, providing some important evidence for the predictive validity of the STEU and STEM.

Table 4.4

Factor Loadings for Structural Analyses of STEU and STEM Item Parcels and DASS-21

	Und	*Man*	*D*	*PA*	*AS*	*S*
STEU Contempt	.306					
STEU Anger	.298					
STEU Regret	.344					
STEU Relief	.343					
STEU Fear	.294					
STEU Surprise	.267					
STEU Joy	.392					
STEU Pride	.268					
STEM Unfairness		.344				
STEM Rudeness		.624				
STEM Anticipation		.274				
STEM Role-Change		.338				
STEM Health Concerns		.273				
STEM Workplace Isolation		.469				
STEM Loneliness		.373				
STEM Parting		.451				
1. Downhearted/blue			.662			
2. Nothing to look forward to			.693			
3. Life meaningless			.651			
4. Worthless as a person			.628			
5. Unenthusiastic			.637			
6. No positive feelings			.661			
7. No initiative to do things			.508			
8. Heart racing				.510		
9. Dry mouth				.457		
10. Breathing difficulty				.627		
11. Trembling				.293		
12. Panic/make fool of self					.641	
13. Close to panic					.654	
14. Scared without reason					.369	
15. Hard to wind down						.732
16. Difficult to relax						.695
17. Nervous energy						.574
18. Agitated						.609
19. Over-react to situations						.444
20. Touchy						.560
21. Intolerant of obstacles						.530

Note. D = Depression; S = Stress; A = Anxiety; PA = Physiological Arousal; SA = Subjective Anxiety.

Table 4.5

Correlations between Factor Scores of Strategic EI and States of Distress

	Und	*Man*	*D*	*PA*	*SA*	*S*
Management (Man)	.72					
Depression (D)	-.12	-.24				
Physiological Arousal (PA)	-.38	-.38	.40			
Subjective Anxiety (SA)	-.54	-.32	.51	.53		
Stress (S)	-.16	-.37	.63	.48	.54	

Note. Only 2 factor inter-correlations (r = -.12 and r = -.16) are not significant at $p < .05$.

Und = Understanding.

3.5 Relationship of Demographic and Life-Style Variables to Strategic EI and the DASS

3.5.1 Relationship between DASS Factors and Criterion Variables

Table 4.6 shows the relationships between factor scores and life-style variables. Both anxiety factors and Stress relate to age (with lower distress for older people). All four distress factors had small relationships with income (with lower distress associated with higher income). There were no gender differences on the DASS latent factor scores, despite prior research indicating that females score more highly on the depression and anxiety subscales (Crawford & Henry, 2003; Henry & Crawford, 2005). This may indicate that the sample is not representative of the population as a whole, an assertion supported by the fact that relatively more women than men volunteered to take part in this study. That gender differences on the DASS are not representative of the population as a whole might also indicate that gender differences in Strategic EI should also be interpreted with caution.

Table 4.6

Correlations between Factor Scores and Life-Style Variables

	Und	*Man*	*D*	*PA*	*SA*	*S*
Age	.16	.15	-.14	-.22**	-.22**	-.22**
Sex (males as reference group)	-.05	-.09	.05	.11	.03	.11
Culture (majority as reference group)	-.18*	-.11	-.07	.10	.08	-.05
Years of education	-.04	-.02	-.04	.00	.02	.03
Annual income	.09	.13	-.19*	-.18*	-.18*	-.18*
Job status	-.09	-.07	-.14	-.02	.01	-.03
No. monthly visits to family	.06	.01	-.03	.12	-.07	-.02
No. monthly visits to friends	-.06	-.05	-.11	-.01	-.02	-.05
Member of community group? (yes = 1)	-.07	-.01	-.04	-.06	.02	-.01
Smoker (non-smokers as reference group)	-.08	-.11	.11	.10	.11	.09
Daily drinker (non-drinkers as reference group)	.13	.15	-.09	-.24**	-.14	-.13

Note. D = Depression; S = Stress; PA = Physiological Arousal; SA = Subjective Anxiety.

*$p < .05$, **$p < .01$

3.5.2 Relationship between Strategic EI Factors and Criterion Variables

There was only one significant correlation between the Strategic EI factors and demographic variables: The cultural majority (those who identified as Australian or Anglo-Celtic) had slightly higher scores on Understanding. However, given that only one of the 22 correlations in Table 4.6 was significant at $p < .05$ (and not at $p < .01$), this seems more likely to be a type I error than a genuine relationship. There were no gender differences on either Understanding or Management, in contrast to results obtained when these constructs are assessed with the MEIS and MSCEIT (Ciarrochi et al, 2000; Day & Carroll, 2004; Kafetsios, 2004; Mayer, et al., 2000b). However, given that gender differences on the DASS also did not follow patterns found in prior research, the lack of gender differences found for Understanding and Management may relate to idiosyncrasies in the current sample, rather than a lack of gender differences on the constructs.

There were also virtually no relationships between the two EI factors and the life-style variables. Even when controlling for work type (part time versus full time), years of education, or gender, there were still no relationships with income or job status. One further finding is that daily drinkers had significantly lower scores on Physiological Arousal than the people who drank less often. Perhaps this relates to the "awareness" aspect of the Physiological Arousal (two items begin with "I was aware of…", and the other two with "I experienced…"), where heavy drinkers are less aware of their symptoms of anxiety.

4. Discussion

Given the restriction of range in this sample, relationships between EI and the DASS scales were supportive for the validity of the STEU and STEM in a non-student sample. However, the lack of relationship between EI and demographic variables indicated a lack of validity in the STEU and STEM. Although reliabilities were lower in the current chapter than in the Chapter 3, STEU and STEM scores were related but differentiable, and predicted state distress. However, there was no relationship between the putative measures of EI and life-style variables. Possible interpretations of these findings and suggested directions for examining EI's relationship to life-style variables are discussed below.

4.1 Comparison of STEU and STEM in Student and Non-student samples

Clearly, both STEU and STEM were not as reliable in this sample as they were for the undergraduate sample in the previous chapter. Given that the STEM's item stimuli were derived from a sample of 60% undergraduate students, and its options from an entirely student sample, this might indicate that the test is not appropriate for non-student samples. However, the STEU does *not* have student-derived content, and suffered greater losses in reliability. Decreases in reliability in the community sample compared to the undergraduate sample corresponded to decreases in variability and difficulty, indicating that reliability decreases might be due to restriction of range. Specifically, means and SDs changed from .61 (*SD* = .13) to .70 (*SD* = .11) for the STEU and from .58 (*SD* = .10) to .60 (*SD* = .08) for the STEM. That is, community volunteers scored significantly higher on both

tests, but the superiority was particularly pronounced for the STEU, where reliability dropped more markedly. An important question is *what causes this restriction of range?* Although the current sample is more highly educated than the Australian norm, the same could be said for Chapter 3's undergraduate sample. Age may be a key issue here, with STEU and STEM scores increasing markedly from young to middle adulthood. However, the use of volunteers who were given feedback on their test scores as their only incentive to participate may have affected scores both through increasing participants' motivation, and resulting in a non-random sample who were particularly interested and particularly adept at EI-related tasks.

The psychology undergraduates in Chapter 3 received course credit as an incentive to participate in the study and were not given any feedback on their results, whereas the community volunteers were aware that they would receive feedback and probably participated primarily in order to receive it. Community volunteers may therefore have been relatively more motivated to achieve high scores. In addition, people who volunteer their time with feedback as their incentive may quite likely have a high degree of interest in the EI construct, and may in fact score more highly on EI tests than people who would *not* volunteer for such a study. For these reasons, the higher scores and lower reliability of the STEU and STEM in the community sample cannot be interpreted as solely due to age and context-familiarity differences. That is, although reliability decreases and scores are higher, this may be due to the non-representative nature of the community volunteers as well as the difference between psychology undergraduates and the general population.

In the current sample, Understanding and Management were clearly differentiable as two separate constructs. There is empirical support for age differentiation of intelligence, where intelligence becomes more differentiated at higher ages, and with higher levels of education (Abad, Colom, Juan-Espinosa, & Garcia, 2003; Anastasi, 1970; Burt, 1954; Spearman, 1927). Given that performance-based EI is considered a form of intelligence, and that the current sample is both older and more highly educated than the sample in Chapter 3, it makes sense that Understanding and Management are more strongly differentiated in this sample. However, the correlation between

Understanding and Management was still quite high (r = .73) indicating that these are very similar abilities.

4.2 Relationship of STEU and STEM to State Distress

As hypothesised, both Understanding and Management factors predicted state distress. The hypothesis that prediction would be stronger for anxiety than for depression or stress was met only for Understanding, where test scores related to anxiety but not to depression or stress. Management predicted all three states of distress approximately equally, and this greater prediction of distress for Management compared to Understanding agreed with hypotheses. Anxiety's stronger relationship to Understanding than Management was consistent with Matthews et al. (submitted) and Bastian et al.'s (2005) results for the MSCEIT. The subjective aspect of anxiety correlated particularly highly with Understanding. Items assessing this subjective experience seemed partly based on situational awareness, which overlaps substantially with what the STEU actually assesses (e.g., items included "I was worried about situations in which..." and "I felt scared *without any good reason*" [italics mine]).

A limitation of the current study was not accounting for intelligence or personality (particularly trait Neuroticism [N] or its facets) in the EI-state distress relationship. Under these circumstances, the relationship of STEM and STEU to state distress might be due to their overlap with intelligence or personality. However, both crystallized (Gc) and fluid intelligence (Gf) relate only minimally to stress reaction, indicating that an intelligence overlap could not be responsible for the STEM-stress relationship (ρ = -.09 for Gc and ρ = -.08 for Gf; Ackerman & Heggestad, 1997). Intelligence relates positively to Strategic EI but generally relates to *higher* rather than lower state anxiety (through the relationship of state anxiety to test anxiety; Zeidner, 1998). Therefore, the STEU and STEM's relationships with lower anxiety would also not be due to overlap with intelligence.

Although Chapter 3 found that neither the STEU nor STEM related to N, they may relate to N's underlying facets, predicting state distress through this pathway. If STEU and STEM scores relate to trait Anxiety (N1), Depression (N3), and Vulnerability (to stress, N6), the relationship with state anxiety, depression, and stress may well be due to this overlap with personality. This would

indicate a relationship with personality, rather than the prediction of lower distress through the correct understanding and effective management of emotional situations (i.e., possibly evidence of poor discriminant validity rather than good predictive validity for measures of performance-based EI). The following chapter examines the STEU and STEM's relationship to the 30 facets of personality in the NEO-PI-R model including trait Anxiety, Depression, and Vulnerability, and thus can provide some evidence of whether the prediction of state distress is due to overlap with personality traits.

4.3 Relationship of STEU and STEM to Positive Life-Style Factors

There was no evidence that the STEM or the STEU related to *any* of the life-style variables. This may constitute a major problem for predictive validity and the utility of EI: If EI measures cannot predict emotion-related criteria then they may not be useful at all. However, the lack of relationship of Strategic EI to the life space criteria examined in this chapter is in fact consistent with Brackett's research on the MSCEIT's relationship to life space criteria (Brackett & Mayer, 2003; Brackett et al., 2004). Although Brackett's studies did find some important predictive validity evidence for the MSCEIT, much of this prediction was from the experiential area rather than the strategic area of EI, or related to criteria not examined in the current chapter. In psychology undergraduate samples, Brackett and Mayer found that the MSCEIT total score did *not* predict tobacco, alcohol, or illegal drug use. Brackett et al. (2004) again found that Strategic EI did not predict tobacco or alcohol use and predicted drug use only in males. Although Brackett et al. found both positive and negative relations with friends were predicted by Strategic EI but the magnitudes were very small (r = .16 and r = -.15, respectively). That is, the fact that the STEU and STEM do not predict lifestyle factors indicates a lack of predictive validity of the Strategic EI construct rather than the STEU and STEM as instruments.

What Strategic EI should predict may also to need to be more carefully specified (e.g., Wilhelm, 2004; Landy, in press, 2005). People at different ages and with different goals may use their emotional intelligence for different purposes, and hence it might predict different things. For example, tobacco use and EI may relate only in people who are trying to quit smoking, where tobacco

use may be a more likely and more salient example of a poor strategy for coping with stress. Similarly, social support may need to be defined differently for different groups (e.g., contact with friends rather than family members is clearly a better way of assessing social support networks for international students studying in a different country from their families). Unfortunately, the current sample was simply too broad to make fine distinctions of this type, although such careful specifications of relevant outcomes for different groups would be a useful direction for further research.

4.4 Conclusions from Chapter 4

Several cautious conclusions regarding the STEU and STEM's validity can be gleaned from this study. Firstly, Management and Understanding *are* differentiable (although very similar), at least in adults with high levels of education. Secondly, there is reasonable evidence of construct validity of STEM and STEU scores, which relate to each other and to the DASS scales. Finally, there is no evidence that STEM and STEU scores relate to broad life-style variables indicative of success in any general way across different groups of people.

Chapter 5

Distinguishing EI from Existing Constructs: A Latent Variable Analysis

1. Introduction to Chapter 5

The current chapter examines the construct validity of Strategic EI (measured by performance-based tests, as in previous chapters) as well as a measure of trait EI (the Schutte Self Report Scale, or SSRI, Schutte et al., 1998). The examination of Strategic EI's relationship to intelligence and personality in Chapter 3 is extended in three ways. Firstly, personality is measured more exhaustively at the facet level of the NEO-PI-R model (Costa & McCrae, 1995) rather than at the broad dimension level only. Secondly, the distinctiveness of Strategic EI from intelligence is examined using multiple indicators of EI, fluid (Gf), and crystallized (Gc) intelligence. Thirdly, the possibility that consensus scoring is partly responsible for the factorial independence of EI from other forms of intelligence is examined.

Several subsidiary issues relating to the validation of the STEU and STEM are also investigated in this multivariate study. These include determining test-retest reliability coefficients for the STEU and STEM and investigating issues of gender and cultural differences relating to possible test bias. The structure of the SSRI measure of trait EI and its independence from personality at the facet level are examined prior to comparing trait and performance-based EI's incremental prediction of alexithymia, stress, life satisfaction, and loneliness. These issues are discussed in further detail in the paragraphs below.

1.1 Psychometric Evaluation of the STEU and STEM: Stability over Time and Group Differences

Two important psychometric issues regarding the STEU and STEM are examined in the current study: test-retest reliability and group differences. As the STEM was developed according to the principles of the Situational Judgment Test (SJT), it may well be multi-dimensional as SJTs often are (Chan & Schmitt, 2002). Similarly, the STEU covers a wide range of emotions (14 different

discrete emotions in all) that might involve differing skills or knowledge bases, again leading to multi-dimensionality of the test. As such, test-retest reliability estimates might be a better indicator of reliability than internal consistency estimates such as Cronbach's alpha.

Gender differences on existing performance-based EI measures have consistently favoured females, particularly for Management (e.g., Ciarrochi et al, 2000; Day & Carroll, 2004; Kafetsios, 2004; Mayer, et al., 2000b). However, Roberts et al. (2001) suggest that the screening sample's demographic characteristics can affect group differences. In their analysis, whites and males did better under expert scoring, where experts were white males. Both expert and consensus screening samples for the STEM are majority female, so that females may do better on the STEM due to the composition of the screening sample, rather than a genuine superiority in their Management. Changing the composition of the screening sample and then assessing whether gender differences remain is a direct test of this idea. If an all-male sample is used to determine scoring weights, and females still perform better under these weights, then female superiority would be genuine rather than artificial. In the current chapter, this analysis will be undertaken for all EI tests that show gender differences, to make sure that such differences are authentic rather than an artefact of scoring procedures. Female superiority on the STEU and STEM is expected, with the difference greater for the STEM, since previous analyses of the MSCEIT have found greater differences for Management. However, it is also expected that female superiority might be exacerbated by the composition of the screening samples.

Given that emotional situations and their solutions are social in nature and possibly dependent on social norms, cultural differences in STEM and STEU scores might be expected, with a possible cultural bias against the non-majority group (who may have slightly different sets of social norms). However, since both the STEU and STEM are highly verbal measures, it is important that analysis of cultural bias take into account English language experience. Although bias based on language is also unacceptable, it is important to tease out these two different issues. It is expected that both the STEU

and STEM will show differences systematically related to English language experience, but will not otherwise be culturally biased.

1.2 Relationship of the STEU and STEM to Personality

Chapter 3 examined the relationship of the STEU and STEM to the five-factor model of personality, finding that both measures related to Agreeableness but not to any of the other four dimensions. This relationship with Agreeableness is consistent with previous research (particularly for the STEM, since Management has shown the strongest relationships to Agreeableness) but the lack of relationship with Openness is not (Brackett & Mayer, 2003; Brackett, et al., 2004; Lopes et al., 2003, 2004; Roberts et al., 2001; Schulte et al., 2004). This chapter expands on the study outlined in Chapter 3 by assessing personality at the facet level of the NEO-PI-R model instead of the broad dimension level only (Costa & McCrae, 1995). The NEO-PI-R personality model is measured with a variation of Goldberg's (2004) International Personality Item Pool (IPIP) version of the NEO-PI-R. There are three primary reasons for this expansion: (1) to assess which facets of Agreeableness relate most strongly to STEU and STEM scores; (2) to assess whether STEU and STEM scores correlate with the emotion-relevant facets of Openness (e.g., Openness to Emotions) rather than the intellect-relevant facets (e.g., Intellect); and (3) to examine whether STEU and STEM scores relate to the facets of Neuroticism, particularly Anxiety, Depression and Vulnerability.[28]

Although Chapter 3 found both STEU and STEM scores related to Agreeableness, only a brief measure of Agreeableness was used, and facet scores were not considered. The relationship of Strategic EI to Agreeableness might conceivably differ for different facets, and the current chapter explores this issue. Similarly, although Chapter 3 found no relationship between Openness and either the STEU or STEM, the brief measure of Openness used in Chapter 3 resembled the more intellectual/cognitive aspects of Openness (i.e., Intellect, Liberalism and Adventurousness) than the

[28] Some facet scores have different labels in Goldberg's IPIP research version of the NEO-PI-R than in Costa and McCrae's (1992) original version. In this chapter, facet scores are referred to by Goldberg's labels, since it is a version of Goldberg's IPIP NEO-PI-R that is used.

emotional/artistic facets of Openness (i.e., Emotion, Artistic Interests and Imagination). It is quite possible that STEU and STEM scores would relate more strongly to the emotional than the intellectual facets of Openness, and this hypothesis is tested in the current study.

Chapter 4 found that the STEU and STEM related to state distress, but did not control for personality. It is thus possible that the STEU and STEM's relationships with *state* anxiety, depression, and stress are due to the overlaps with *traits* of Anxiety, Depression, and Vulnerability (to stress). Ideally, the STEU and STEM should relate to state distress through effective understanding and management of emotional situations, rather than because test scores overlap with personality (i.e., Neuroticism [N], or negative emotionality). Although Chapter 3 found that STEU and STEM scores were unrelated to N, it is possible that they still relate to the underlying facets of N. Particularly relevant for interpreting the results reported in Chapter 4 are the relation of STEU and STEM scores to the NEO-PI-R's to Anxiety, Depression, and Vulnerability facets.

It is hypothesised that both STEU and STEM should be relatively independent from all facets of the NEO-PI-R ($r < .30$), as they are measuring a form of cognitive ability rather than an aspect of personality. However, scores are expected to relate moderately to Agreeableness and to the emotion-related facets of Openness. STEU and STEM scores are *not* expected to relate meaningfully to Anxiety, Depression and Vulnerability

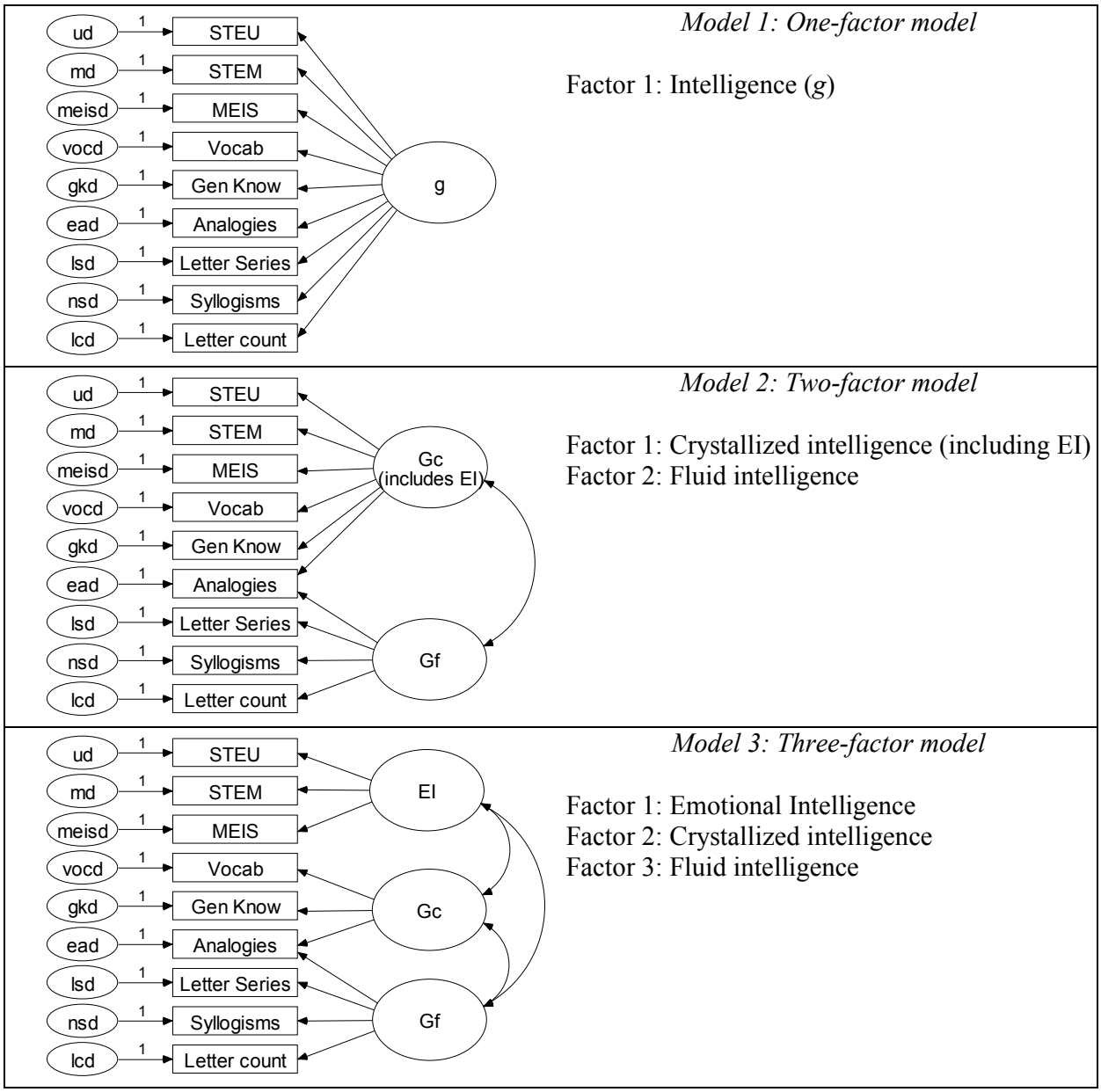
1.3 Distinction of Strategic EI from Fluid and Crystallized Intelligence

Confirmatory factor analysis with multiple indicators of Strategic EI, Gf, and Gc will be used to examine the relationships between Strategic EI, Gf, and Gc at the construct rather than at individual test level (cf. Chapter 3). In the current study, three indicators for each construct are used. Strategic EI is represented by short forms of the STEU and STEM and an Understanding composite from the MEIS. Gf is represented by letter series, letter counting, and syllogistic reasoning tasks. Gc is represented by vocabulary, analogies and general knowledge tasks. Nested models test whether a one-factor (*g*), two-factor (Gf and a Gc component that includes EI), or three-factor solution (Gf, Gc, and EI) better fit the data, allowing a test of whether Strategic EI is independent from Gc. This nested

models design is shown in Figure 5.1. The paragraphs below outline in more detail how the measures were chosen and how this analysis was expanded to test whether consensus scoring may act as a method effect in determining EI's structural independence from intelligence.

Figure 5.1

Planned Analysis for Testing Relationships between Fluid, Crystallized and Emotional Intelligence: Comparison of Three Nested Models



Note. These analyses are run twice: once when EI tests are consensus scored and Gf and Gc are scored conventionally and again when EI, Gf, and Gc are all consensus scored (see text for details).

1.3.1 Ratings-Based Response Format and Consensus Scoring as Method Effects

1.3.1.1 Controlling for Ratings-Based Response Format

As previously discussed, the MSCEIT test battery is virtually the only comprehensive performance-based EI test battery commonly used in both research and commercial settings. However, its tests have both idiosyncratic ratings-based response formats and unusual consensus-based scoring protocol for an intelligence test battery. Chapter 3 found some support for a ratings-based response format: The ratings-based STEM correlated more highly than the multiple-choice STEM with a ratings-based EI test (Stories), and less highly with a multiple-choice EI test (the STEU). The ratings-based STEM also correlated less strongly than the multiple-choice STEM with an intelligence measure (Vocabulary), although not significantly so.

Collectively, these results are supportive of a possible method effect of response format, where multiple-choice tests correlate more highly than ratings-based tests with intelligence (possibly since intelligence is not typically measured with ratings-based items). For this reason, all EI indicators were chosen to be multiple-choice rather than ratings-based, such that the distinction between EI and intelligence cannot be due to response format. However, this design issue meant that additional EI indicators (other than the STEU and STEM) represented the construct of Emotional Understanding only and not Emotion Management, since all Management tests from the MEIS and MSCEIT are in ratings-based format.

1.3.1.2 Exploring Consensus Scoring as a Method Effect

Davies et al. (1998) suggested that consensus scoring may act as a method effect, where EI forms a distinct factor due to the distinctive scoring used in EI tests, rather than the distinctiveness of the construct. To present knowledge, this has not yet been empirically examined and thus the

investigation of this issue is an important focus of the current chapter.[29] In this chapter, Gf and Gc tests are scored by consensus (in addition to conventional methods) to test whether consensus scoring constitutes a possible confound in distinguishing between EI and other cognitive abilities.[30] The analyses outlined in Figure 5.1 tests a one-factor model (where Gf, Gc, and EI all form a general cognitive ability factor) against a two-factor model (where EI forms part of a Gc factor, but Gf is distinct) against a three-factor model (where Gf, Gc, and EI are distinct factors). These three analyses will be run twice: (1) with Gf and Gc indicators conventionally-scored and EI tests scored by consensus; (2) with Gf, Gc, and EI tests *all* scored by consensus. These analyses allow the relationships between EI and intelligence constructs to be compared with and without the potential effect of consensus scoring.[31] It is expected that there will be a method effect of consensus scoring, such that EI and Gc will be more strongly related when both are scored by consensus than when only EI is consensus-scored.

1.3.2 Factor Analysis of Strategic EI, Gf, and Gc

It is expected that Strategic EI will be much more strongly related to Gc than Gf, and hence distinction from Gc (rather than Gf) is necessary for EI to be a separate type of intelligence from those already well-known and well-researched. This hypothesised relationship of EI with Gc (rather than Gf) forms the conceptual basis for the nested models design shown in Figure 5.1. There are at least three grounds for the assertion that EI relates to Gc rather than Gf. Firstly, as discussed in Chapter 1, MEIS and MSCEIT tests of EI relate much more strongly to Gc than to Gf. Within the four-branch structure of EI, Strategic EI (and especially Understanding) relates much more strongly

---

[29] Although one of the three studies in Davies et al. (1998) was designed to examine consensus-scoring as a method effect, the exclusion of several unreliable consensus-scored tests meant that this paper was unable to empirically examine this issue as planned.

[30] Note that scoring intelligence tests by consensus was undertaken for experimental purposes only, and was certainly not meant to imply that this is an appropriate alternative to conventional forms of scoring.

[31] Ideally EI tests would also be scored both veridically and by consensus, but veridical scoring is only possible for the STEU (as Chapter 3 found that the reliability of the STEM was virtually zero, and MEIS tests also cannot be scored veridically).

to Gc than do other branches. Secondly, it is fairly obvious that tests which involve processing relatively large amounts of text will relate to verbal ability. In this case, relationships between EI and Gc represent extraneous or irrelevant variance in EI test scores due to verbal ability. However, Chapter 3 found that EI-EI correlations did not diminish much when vocabulary-related variation was accounted for, indicating that this may not be a major concern for the STEU and STEM. Thirdly, theoretical descriptions of EI (and again, particularly Strategic EI and particularly Understanding) point to its similarity to Gc. Mayer et al. (2001) suggest that EI develops with age and experience and that an acquired emotional knowledge base underlies emotional understanding abilities. Thus age trends and developmental processes mirror Gc (increases across the life span and development of knowledge through acculturation). Other researchers have suggested that EI may form a component or primary mental ability of Gc for these reasons (Davies et al., 1998; Roberts et al., 2001). Although this study examines only Strategic EI and therefore covers half the content domain of emotional intelligence, it is at this highly verbal and abstract level where distinction from existing intelligences may be more problematic (and is thus a stricter test of the discriminant validity of performance-based EI).

Thus the structural investigation of EI, Gf, and Gc has two purposes. Firstly, the analysis tests whether Strategic EI may be distinguished from Gc. Secondly, the analysis tests whether this distinction is stronger when Gf and Gc tests are also consensus-scored, and thus whether consensus scoring acts as a method factor. If there is evidence for this method factor, there are important implications for interpreting existing research that is based on consensus-scored EI tests. As Davies et al. suggested, EI's emergence as a separate factor from other cognitive abilities may be due to method-related variance rather than a construct-related variance. More specifically, the distinction between EI and intelligence may be wholly or partly due to this consensus-scoring method factor.

1.4 Structural Analysis of Trait EI: The Schutte Self-Report Scale (SSRI)

The current study plans to evaluate structural validity of trait EI in respect to one instrument in particular: the 33-item SSRI from Schutte et al. (1998).[32] Item content was drawn from a precursor to the hierarchical four-branch model of EI consisting of three broad dimensions: appraisal/expression of emotion, utilisation of emotions in problem solving, and emotion regulation (Salovey & Mayer, 1990). Although content was taken from these three related domains, the scale was constructed to be uni-dimensional. Despite this, Petrides and Furnham (2000) and Saklofske et al. (2003) both found very similar four-factor structures, where factors represented: (1) Optimism/Mood Regulation; (2) Appraisal; (3) Utilisation of Emotions and (4) Social Skills. However, Austin et al. (2004) found the fourth factor (Social Skills) was poorly defined, and proposed a three-factor model instead, which was replicated in Austin, Saklofske, and Egan (2005). Gignac, Palmer, Manocha, and Stough (2005) also did not find a Social Skills factor, although their model still contained four factors since Appraisal was separated into Self-Appraisal and Other-Appraisal factors. Given the level of disagreement, one aim of the current study is to compare one-factor, three-factor, and four-factor structures to see which model is the most parsimonious. Given the more recent replications of the three-factor model with more sophisticated statistical techniques, the three factor model is expected to provide the most parsimonious fit.

1.5 Distinction of Trait EI from the NEO-PI-R Model of Personality

Many items measuring trait EI are conceptually and semantically similar to personality items, making this the main challenge to the discriminant validity of trait EI (MacCann et al., 2003). Although Schutte et al. (1998) found the SSRI had only small ($r < .30$) relationships with all five-factor dimensions except for Openness ($r = .54$), this was with a sample size of 23 and later results do not confirm this finding (Austin et al., 2005; Saklofske et al., 2003). In fact, both Austin et al. and

---

[32] Note that this scale has been referred to by a number of different acronyms in different papers, but will be referred to as the SSRI throughout this chapter

Saklofske et al. found the SSRI related less than $r = .30$ with Openness, but related strongly to Extraversion ($r = .45$ and $r = .51$, respectively). Austin et al. also found strong relationships with Agreeableness ($r = .58$).

These two studies are typical of most studies examining trait EI's relationship to the five factor model personality, in that their examinations stop at the broad dimension rather than facet level (e.g., Austin, 2004; Austin et al., 2004; Furnham & Petrides, 2003; Petrides & Furnham, 2003). However, Petrides and Furnham (2001) provide an exception to this rule, and Petrides and Furnham (2003) measured personality at the facet level but did not report facet-level information. Examination at the narrower facet level as well as the broad dimension level is necessary as components of trait EI might be replicating narrow facets (rather than broad dimensions) of personality. For example, both O'Conner and Little (2003) and Petrides and Furnham (2001) found very strong relationships between the EQ-i and trait anxiety (a facet of Neuroticism). In the current study, relationships between the SSRI and personality will be examined at the facet level of personality.

As well as measuring personality at a narrower level, measuring EI at a narrower level would also be informative in assessing the relationship between personality and trait EI. Specifically, the current study plans to look at the relationship between the NEO-PI-R facets and each factor of the SSRI. Saklofske et al. (2003) examined the relationship between each factor of the SSRI and the five factors of personality, finding that the SSRI-Extraversion relationship was accounted for by only two of the four factors of the SSRI: Optimism/Mood Regulation and Social Skills. Saklofske et al.'s analysis also showed that Optimism/Mood Regulation correlated higher than $r = .45$ with *three* of the five factors of personality (Conscientiousness, Extraversion, and Neuroticism). The current study expands on Saklofske et al.'s analysis by examining personality relationships with SSRI factors for all 30 facet scores of the NEO-PI-R model of personality (rather than for the five broad dimensions only). It is expected that the Optimism/Mood Regulation factor of the SSRI will relate to Extraversion, Neuroticism, and Conscientiousness. This factor might relate particularly strongly to

facets conceptually similar to *Optimism*, such as Cheerfulness (E6) and Depression (N3), or to *Mood Regulation*, such as Self-Efficacy (C1), Self-Discipline (C5), or Assertiveness (E3).

1.6 Incremental Validity of EI: A Comparison of Performance-Based Strategic EI and Trait EI

An important difference between EI and other measures of intelligence is EI's theoretical link with criterion measures relating to emotional well-being. Both performance-based and trait EI have been linked with *life satisfaction* (Austin et al., 2005; Ciarrochi et al., 2000; Mayer et al., 2000b, see also Chapter 3 of this volume); *loneliness* (Engelberg & Sjoberg, 2004); *stress* (Ciarrochi et al. 2002; Salovey et al., 2002; Woitaszewski & Aalsma, 2004; also Chapter 4 of this volume); and *alexithymia* (Austin et al., 2005; Barchard & Hakstian, 2004; Dawda & Hart, 2000; Parker et al., 2001; Warwick & Nettelbeck, 2004; also Chapter 3 of this volume).

Given the diverse operationalisation of both EI and measures of well-being, results are difficult to clearly summarise. However, self-report measures of well-being are more strongly related to self-report EI than performance-based measures of EI. Only one study has controlled for *both* personality *and* intelligence in the relationship between EI and well-being measures (Bastian et al., 2005). This chapter will follow that lead, examining the relationship of EI to well-being while controlling for personality and intelligence. It is hypothesised that the different predictive strength of trait EI compared to performance-based EI will disappear once personality is accounted for. It is further hypothesised that EI (both trait and performance-based) will provide some incremental prediction of loneliness, life satisfaction, and stress once personality and intelligence are accounted for.

1.7 Summary of Hypotheses

Hypothesis 1: Test-retest Reliability of the STEU and STEM. Test-retest reliability of the STEU and STEM is expected to be higher than internal consistency reliability, indicating that the tests may be multi-dimensional.

Hypothesis 2: Group Differences on the STEU and STEM. Female superiority is expected for both the STEU and STEM, with differences greater for the STEM. It is further expected that such

gender differences will decrease when accounting for the female-majority composition of the screening samples used to determined consensus weights. In addition, no significant differences due to cultural identification on the STEU and STEM are expected when accounting for English speaking experience.

Hypothesis 3: Distinction of Strategic Performance-Based EI from Gf and Gc. It is expected that Strategic EI will be a separate construct from both Gf and Gc, although far more strongly related to Gc than Gf, since it is verbal and knowledge-based. It is further expected that Strategic EI and Gc will relate more strongly when Gc tests are consensus scored, providing evidence of a method effect of scoring.

Hypothesis 4: Factor Structure of the SSRI. Given recent evidence and the content domain sampled, it is hypothesised that a three-factor structure of the SSRI will be superior to a four-factor structure, and that both of these will be better than a one-factor structure. This will be tested with confirmatory factor analyses.

Hypothesis 5: Distinction of Trait EI from the NEO-PI-R Model of Personality. SSRI and TAS-20 components are expected to show relative independence from the NEO's 30 facets, correlating less than $r = .50$ (the cut-off for a large correlation from Cohen, 1988) with all individual facets.

Hypothesis 6: Relationship of Strategic EI to the NEO-PI-R Model of Personality. Scores on the STEU, STEM and MEIS composite will be compared to facet scores on the NEO-PI-R model of personality. It is expected that EI scores will be relatively independent from personality facets, correlating at less than $r = .30$ (the cut-off for a small correlation from Cohen, 1988). However, non-trivial correlations with facets of Agreeableness, and with the emotion-related facets of Openness are expected.

Hypothesis 7: Relationships of Trait and Ability EI to Emotional Well-Being. It is expected that both trait and ability based EI will relate to states of loneliness, life satisfaction and stress and to alexithymia (particularly the externally-oriented-thinking component), but that this relationship will

be far stronger for trait EI before personality is accounted for. It is further expected that relationships of trait EI with such state measures will decrease markedly once personality is accounted for.

2. Method

2.1 Participants

181 third year undergraduate psychology students (136 female[33]) undertaking a course on psychological assessment at the University of Sydney completed tests as part of their class work over five tutorial sessions. Ages ranged from 19 to 59 years (M = 21.9, SD = 3.7 median = 21). 36 participants learnt English as a second language, and had been speaking English between 4.5 and 21 years (median = 13; age at which English was learnt ranged from 2 to 23). 109 participants identified with Australian or Anglo-Celtic (e.g., British, New Zealander) cultures (95 as *only* with these cultures and no others). 46 identified with Asian cultures, 29 with European, 6 with Middle-Eastern, 2 with South American, and 1 with Indigenous Australia. In analyses comparing cultural groups, cultural differences are considered for the three largest groups only: Australian/Anglo-Celtic, Asian, and European (i.e., cultural groups were defined at a very broad and basic level by self-identification only).

2.2 Test Battery

2.2.1 Emotional Intelligence Tests

Participants completed four measures of performance-based EI and one measure of trait EI. All performance-based tests were from the strategic area of EI (three representing Understanding and one Management), and all were in multiple-choice format (so as to avoid possible method effects in differentiating EI from Gf and Gc).

[33] One person in this session did not report any demographic information, including gender, age, and cultural identification. There were also several (27) participants who completed other sessions but not the first one, and therefore no demographic information about them is available.

Situational Test of Emotional Understanding (STEU). Participants completed the 14 personal-life-context and 14 context-less items of the STEU, described in earlier chapters and in Appendix 2.1. In the current study, the STEU is scored by consensus.

Situational Test of Emotion Management (STEM). Participants completed 7 sadness-management, 7 fear-management and 7 anger-management items. Items were selected by examining the data from Chapter 3 and selecting those items with the highest reliability under both classical test theory and item response theory. A more detailed justification for the choice of these 21 items is given in Appendix 5.1. In the current study, the STEM is scored by consensus.

Blends test from the Multi-factor Emotional Intelligence Scale (MEIS, Mayer, Salovey, and Caruso, 1997). This 9 item, multiple-choice test can be scored according to expert or consensus criterion, and measures emotional understanding.[34] Tests from the Understanding branch of the MEIS were chosen because they are the only ones that are in a multiple-choice format (all other tests are in a rating-scale format). In the current study, Blends is scored by consensus. A sample item follows: *Optimism most closely combines which two emotions? (a) pleasure and anticipation, (b) acceptance and joy, (c) surprise and joy, (d) pleasure and joy.*

Changes test from the Multi-factor Emotional Intelligence Scale (MEIS, Mayer et al., 1997). This 9 item, multiple-choice test can be scored according to expert or consensus criterion, and measures emotional understanding. In the current study, Changes is scored by consensus. A typical item follows: *Feeling livelier and livelier results in? (a) joy, (b) loving, (c) surprise, (d) excitement.*

Schutte Self-Report Scale (SSRI; Schutte et al., 1998). This 33-item scale measures emotional intelligence. Participants rated their agreement with each statement of a 6-point scale from "Strongly Disagree" to "Strongly Agree".[35] E.g., *I know what other people are feeling just by looking at them.*

---

[34] Both the Blends and Changes tests consist of 8 items and one example item. In this study, the example items from the MEIS were included as part of the tests, making the test 9 items rather than 8 items long.

[35] In previous research, the both the SSRI and TAS-20 have used a 5-point rating scale, and the DASS stress scale a 4-point scale, rather than a 6-point scales used in the current study. However, existing research indicates that the

2.2.2 Criterion Measures of Emotional Well-Being

Toronto Alexithymia Scale (TAS-20; Bagby, Taylor, & Parker, 1994a). This 20-item scale measures three aspects of the clinical condition alexithymia (literally "no words for emotions"): difficulty identifying feelings (DDF), difficulty describing feelings (DDF), and an externally-oriented thinking style (EOT). Participants rated their agreement with each item on a 6-point scale from "Strongly Disagree" to "Strongly Agree".[27] Example items are given in the method section of Chapter 3. In Chapter 3, EOT related to both the STEU and STEM, whereas DDF and DIF related to the STEM only, and only at high levels of vocabulary.

UCLA Loneliness Scale (short-form, Hays & DiMatteo, 1987). This 8-item scale is the short-form measure of loneliness adapted from Russell, Peplau, and Cutrona's scale (1980). Participants were asked to think about how they felt at the current point in time, and then rate their degree of agreement with each item on a 7-point scale from "Strongly Agree" to "Strongly Disagree". E.g., *I feel isolated from others.*

Satisfaction with Life Scale (SWLS; Pavot & Diener, 1993). This 5-item scale is a measure of subjective satisfaction with life. Participants were asked to think about how they felt at the current point in time and rate their degree of agreement with the five items on a 7-point scale from "Strongly Agree" to "Strongly Disagree". The SWLS may be broken into current and retrospective components (McDonald, 1999) and Chapter 3 found that the STEM related to the retrospective component of the SWLS only. An example item is given in the method section of Chapter 3.

Stress Scale from the Depression Anxiety Stress Scale (DASS; Lovibond & Lovibond, 1995). The stress items were taken from the short form of the DASS, and participants were asked to rate the frequency with which they had felt stress-related symptoms in the past week on a 6-point scale from

---

mid-point in odd numbered scales may be misunderstood as an "I don't know" or "Not applicable" response, rather than a point on the scale (Raaijmakers, van Hoof, Hart, Verbogt, & Vollebergh, 2000). For this reason, 6-point scales were used for the SSRI, TAS-20 and Stress scale of the DASS. However, the 7-point scales were retained for the Satisfaction with Life Scale, for continuity with Chapter 3 and Palmer et al. (2002), and due to the difficulty of transforming 7 semantic labels into 6 equivalent ones. The loneliness scale was also rated on a 7-point scale.

"Never" to "Always". [27] Chapter 4 found that the Stress scale related to STEM scores but not to STEU scores. An example item is given in the method section of Chapter 4.

2.2.3 Personality

International Personality Item Pool NEO Test of Personality (IPIP-NEO, short form; Johnson, 2004). This 120-item instrument was based on the NEO-PI-R model of personality (see e.g., Costa & McCrae, 1995), where five broad dimensions of personality (Openness to Experience [O], Conscientiousness [C], Extraversion [E], Agreeableness [A], and Neuroticism [N]) are each composed of six narrow facets of personality. For example, N is composed of Anxiety (N1), Angry Hostility (N2), Depression (N3), Self-Consciousness (N4), Impulsiveness (N5), and Vulnerability (N6). Johnson (1991) created the short form of the IPIP-NEO-PI-R by selecting the four most reliable items for each facet from the complete IPIP-NEO PI-R (Goldberg, 2004). In each item, participants rate the accuracy with which a particular statement (e.g., "get back at others") describes them on a 6-point scale from very inaccurate (1) to very accurate (6). Information on the psychometric properties of the short-form scale is available in Johnson (1991). This scale was chosen to allow a measure of each of the facets of personality within a minimum testing time period.

2.2.4 Cognitive Ability Tests

Participants completed six markers of cognitive ability. All but one of these tests (Nonsense Syllogisms) were adapted from Stankov's (1997) Gf/Gc Quickie Test Battery (Nonsense Syllogisms came from French, Ekstrom, and Price, 1963). Three tests measured Gf, and three Gc. Short forms were used for five of these six tests (all but Vocabulary) with items selected for their high reliability in data sets from Kleitman and Stankov (2005) or MacCann et al. (2004).

Vocabulary Test (Stankov, 1997). This 18-item test is a marker of Gc. Participants must choose from five alternatives which word is the most similar to a target word. E.g., *Revolve: 1. A gun, 2. Uprising, 3. Turn around, 4. Grow, 5. Decide*.

Esoteric Analogies (Stankov, 1997). An 18-item form of Stankov's (1997) original 24-item test was created by selecting the most reliable items from Kleitman & Stankov (2005). This test is

primarily a marker of Gc, although previous research has found cross-loadings on Gf factors also (e.g., Roberts & Stankov, 1999). Participants were given a relationship between two entities and asked to apply this relationship to a third entity, where one of four choices was related to the third in the same way that the first and second were related. Eg: *FIRE is to HOT as ICE is to (a) POLE, (b) COLD (c) CREAM (d) WHITE*.

General Knowledge (Stankov, 1997). This 10-item test indexes Gc, and is a short form of Stankov's (1997) General Knowledge Test used in Kleitman and Stankov (2005). Participants are asked a factual question and must choose the correct answer from five alternatives. E.g., *What is the outermost planet in the solar system? (a) Venus, (b) Saturn, (c) Pluto, (d) Uranus, (e) Earth*.

Letter Series (Stankov, 1997). This 12-item test indexes Gf, and is a short form of the Stankov's (1997) 15-item test, created by selecting the most reliable items from MacCann et al. (2004). Participants were presented with a several sequences of letters and instructed to determine which letter would occur next in each sequence. E.g., *J K L M N O P Q ?*

Nonsense Syllogisms (French et al., 1963). This 12-item test is a short form of the French et al. version, and is a marker of Gf (the 12 items with the highest item-total correlation were chosen from Kleitman and Stankov, 2005). Participants are presented with a syllogism with two nonsense premises and a conclusion based on these, and asked to determine whether the conclusion represents good reasoning or poor reasoning, assuming that the premises are true. E.g., *Some dogs are seals. Some seals bark. THEREFORE, some dogs bark*.

Letter Counting (Stankov, 1997). This 12-item test is a short form of the test given in Stankov (1997). The 12 most reliable items were selected from MacCann et al. (2004). This test is a measure of working memory, which has been strongly implicated in fluid reasoning abilities and correlates very highly with Gf tasks (Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Suess, 2005). Some researchers propose that this relationship is strong enough that working memory can be used as a measure of Gf (Kyllonen & Christal, 1990). Participants were presented between 8 and 10

combinations of the letters R, S, and T displayed serially on the computer 1 second apart, and then had to answer how many Rs Ss and Ts there were in the combination.

2.3 Design and Procedure

Participants completed all test as part of the undergraduate class work over five occasions. All tests were self-paced, in groups of approximately 20. After completing each session, participants were debriefed, and given detailed information about what each construct measured (since this was a class exercise designed for educational as well as research purposes). Data was identified by participants' student ID code and later matched by this, but was kept confidential at all times. Tests were administered over a 13 week period in five sessions, outlined in Table 5.1. This study was approved by the Sydney University Human Research Ethics Committee.

Most participants completed tests on Macintosh computers (due to a shortage of computers, 34 students completed paper and pencil versions on the first occasion, 17 on the second occasion, 22 on the third occasion and none on the fourth or fifth occasion).[36] Due to absences and course withdrawals there are different numbers of participants on different occasions, such that the list-wise *N* differs for different analyses.

---

[36] Pre-analysis compared test scores for computer-administered versus paper-administered tests for the 48 scores where some participants did paper tests (i.e., 3 performance-based EI scores, 4 SSRI components, 3 TAS-20 components, Life Satisfaction, Loneliness, Stress, all cognitive ability but letter counting, and the 30 personality facets). Significant mean differences between computer versus paper administration were found for four scores: (1) the STEU (M = .61, SD = .08 for computer, M = .64, SD = .05 for paper, t_{177} = 2.644, p = .009); (2) Life Satisfaction (M = 21.29, SD = 5.15 for computer, M = 23.76, SD = 6.30 for paper, t_{172} = 2.406, p = . 0177); (3) the Utilising Emotions scale of the SSRI (M = 22.73, SD = 3.07 for computer, M = 23.91, SD = 3.03 for paper, t_{171} = 2.025, p = . 044); and (4) the Vulnerability facet of Neuroticism (M = 14.97, SD = 3.87 for computer, M = 17.53, SD = 3.86 for paper, t_{155} = 2.572, p = .011). These differences were not large, and were not felt to be a problem for the study.

Table 5.1

Test Administration Order over Thirteen Weeks and Five Testing Sessions

Session and Week No.	*Order in Which Tests were Taken*
Week 1 (Session 1)	*EI Measures and Well-Being Measures* (1) STEU; (2) STEM; (3) Blends; (4) Progressions; (5) UCLA Loneliness + SWLS (SWLS items embedded within loneliness scale); (6) DASS Stress; (7) SSRI + TAS-20 (TAS-20 items embedded within SSRI).
Week 3 (Session 2)	*All Cognitive Ability Tests but Letter Counting (Counter-Balanced across Two Ordering Conditions with Participants Randomly Assigned to a Condition)* *Condition 1:* (1) Esoteric Analogies; (2) General Knowledge; (3) Letter Series; (4) Nonsense Syllogisms; and (5) Vocabulary. *Condition 2:* (1) Vocabulary; (2) Nonsense Syllogisms; (3) Letter Series; (4) General Knowledge; and (5) Esoteric Analogies.
Week 4 (Session 4)	*Personality only* Short form of the IPIP-NEO-PI-R
Week 11 (Session 4)	*STEU and STEM (for the second time, ten weeks after the first administration)* (1) STEU; (2) STEM
Week 13 (Session 5)	*Additional Cognitive Ability Test* Letter Counting

2.4 Analyses

Analysis 1: Descriptive Statistics: Test-Retest Reliability and Group Differences for EI. Descriptive statistics for all variables are presented to make sure that the tests are behaving appropriately. The STEU, STEM, and MEIS Understanding composite were analysed in further detail, including a comparison of test-retest reliability with internal consistency reliability, as well as gender and cultural differences.

Analysis 2: Nested Models Confirmatory Factor Analysis of Gf, Gc, and EI Indicators. The set of nested confirmatory factor analysis models presented in Figure 5.1 were run twice: (1) with EI indicators consensus-scored and Gf and Gc indicators dichotomously scored, as by convention; and (2) with all indicators (Gf, Gc, and EI) consensus-scored.

Analysis 3: Exploratory and Confirmatory Factor Analysis of the SSRI Items. The structure of the SSRI was investigated using exploratory factor analysis. Solutions were then checked using confirmatory factor analysis. It was expected that a three-factor solution will be superior to a four-factor solution and that both of these will be superior to a one-factor solution.

Analysis 4: Correlations between EI and NEO-PI-R Personality Facets. Correlations between the 30 personality facets and EI were calculated for each measure of EI used in this study, both the performance-based Strategic EI tests (STEU, STEM, and MEIS Understanding composite) and the SSRI (separately for each factor).

Analysis 5: Regression Models Predicting Emotional Well-Being from Performance-Based and Trait EI Controlling for Personality and Intelligence. Two regression models were run predicting each of the criterion variables: Loneliness, Stress, Life Satisfaction, and three aspects of alexithymia (difficulty identifying feelings, difficulty describing feelings, and externally oriented thinking). For each criterion variable, a simple regression and a three step hierarchical regression were run separately for each EI predictor (i.e., the STEU, STEM, MEIS, and SSRI scores). The simple regressions predicted criteria from EI alone. The hierarchical regressions consisted of three steps. In the first step, Gf and Gc composite scores were entered. In the second step, the five factor dimensions of personality were added (i.e., Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism). In the third step, the EI variable was entered. This set of analyses allowed a comparison of the incremental validity of trait and performance-based EI after personality and intelligence were accounted for.

3. Results

3.1 Reliability and Descriptive Statistics

3.1.1 Reliability and Descriptive Statistics for Performance-Based EI

Reliabilities and descriptive statistics for the performance-based EI tests are shown in Table 5.2. All tests are scored by consensus.[37] Internal consistency estimates are generally low but acceptable for correlational research. As the initial reliability estimate for the STEU was particularly low (Cronbach's alpha = .42), unreliable items were deleted to increase internal consistency. A 15-item version of the STEU was created, using half the sample to calculate item-total correlations (and delete items where these were low) and the other half to calculate internal consistency. It was not possible to increase the reliability of the STEM by deleting items in this way. As the Blends and Changes tests contained only 9 items each, these 18 items were combined to form a more reliable measure, rather than deleting items from each test.[38] Further analyses use the composite Blends and Changes score.

---

[37] STEU and STEM scores were calculated using Chapter 3 consensus weights, Chapter 4 consensus weights and current sample consensus weights, and the most reliable set of weights used (Chapter 4 consensus weights for the STEU, and current sample consensus weights for the STEM). MEIS scores were calculated using New Hampshire consensus weights (i.e., the weights used in Mayer et al., 2000b and Roberts et al., 2001) and current sample consensus weights and only the most reliable set of weights (current sample consensus) included.

[38] Nested model analyses of Blends and Changes items comparing a one-factor solution (Blends + Changes) and two-factor (Blends versus Changes) solutions found *no* significant improvement in fit when Blends and Changes items defined separate latent factors, suggesting that these can be validly combined ($\Delta\chi^2(1) = 2.451$, $p = .1174$).

Table 5.2

Reliabilities, Descriptive Statistics and Group Differences for Strategic EI Tests

	Descriptive Statistics and Reliability				*Group Differences*				
	Reliability		*Descriptive Statistics*		*Sex*	*Culture*			
						No control		*English controlled*	
	Alpha	*Test-Retest*	*Mean*	*SD*	*male – female*	*majority – other*	*Asia – Europe*	*majority – other*	*Asia – Europe*
STEU (*N*=179; 15 items)	.55	.66	.57	.08	-.19	.21**	-.11	.08	-.06
STEM (*N* = 176*;* 21 items)	.50	.55	.65	.07	-.76**	.16*	-.01	.01	.04
Blends (*N* = 179*;* 9 items)	.46	-	.58	.09	.24	.20**	.06	.08	.11
Changes (*N* = 179*;* 9 items)	.38	-	.58	.08	-.06	.21**	-.07	.05	-.02
Blends + Changes (18 items)	.56	-	.58	.07	.13	.26**	.00	.09	.06

Note. Cultural differences compare three groups (Australian/Anglo-Celtic [majority]; Asian; and European). Effect sizes for culture were calculated as the contrast (derived from standard-form coefficients) divided by the pooled SD. Age at which participants first spoke English was used as a covariate in a series of ANCOVAs predicting EI scores from cultural identification. Significance of group differences was determined from the *t* statistic.
* $p < .05$, ** $p < .01$

Females scored more highly than males on the STEM, as hypothesised. However, no female advantage was found for the Understanding tests, either the STEU or the MEIS composite. Given that females form a majority of the consensus group this female advantage may be due to the composition of the weighting sample, as Roberts et al. (2001) point out. However, when weights were determined using a male-only sample, the female advantage remained and, in fact, was slightly larger (d = -.80, t_{173} = -4.655, $p < .001$). Female superiority on the STEM does appear to be a genuine finding rather than an artefact of scoring. There are significant differences between cultural groups on all performance-based EI scores with the majority group scoring significantly more highly. However,

when English language experience was controlled for these differences disappeared, suggesting that it is the verbal load rather than the test content that is the cause of these differences.[39]

Two possible reasons for the low reliabilities obtained for the STEM and STEU are: (1) these are multi-dimensional constructs which have low internal consistency (i.e., support for hypothesis 1); and (2) the relative restriction of range in a senior psychology course, where the participants are studying psychological assessment. Test-retest reliability of the STEU and STEM at 10 weeks was slightly higher than internal consistency reliability, with reliability estimates of .66 and .55 for the STEU and STEM, respectively. The fact that test-retest reliability estimates were higher than internal consistency estimates supports the idea that the constructs measured might be multi-dimensional (see Chan and Schmitt, 2002's suggestions that this is a common problem in SJTs).

However, restriction of range may also be an issue here. To test whether there were restriction of range issues, means on the common items of the STEU and STEM (14 items of the short form STEU and 19 items of the STEM) were compared to samples from Chapters 3 and 4. Since this sample also completed the same vocabulary test as participants in Chapter 3's study, mean scores on the vocabulary test are also compared to provide a rough index of intelligence differences between samples. Comparisons are shown in Table 5.3, and it is immediately obvious that the current sample is a much more select group.

---

[39] The Asia/Europe distinction might also be an international versus local student distinction (rather than a cultural distinction), since the vast majority of international students in Australian Universities are from Asian countries, whereas students identifying with European cultures are local rather than international students. In 2004, 82% of international University students in Australia were from Asia, and only 8% from Europe (International Education Network, June 2005) and in 2002, overseas students made up 22% undergraduate students (Australian Bureau of Statistics, July 2005). Therefore, the best estimate is that 18% of students in the current sample are international students from Asian countries).

Table 5.3

Comparison of Reliability and Descriptive Statistics across Samples from Chapters 3, 4 and 5 for Vocabulary and the Common Items of the STEU and STEM

	N	*Alpha*	*Mean*	*SD*	*d*
STEU (14 common items)					
Chapter 3	112	.51	.56	.09	1.03
Chapter 4	149	.48	.61	.09	0.42
Chapter 5	176	.48	.64	.08	-
STEM (19 common items)					
Chapter 3	200	.68	.55	.11	0.60
Chapter 4	149	.24	.59	.06	0.26
Chapter 5	179	.56	.61	.08	
Vocabulary Test (18 items)					
Chapter 3	201	.79	.49	.20	0.39
Chapter 5	164	.65	.56	.15	-

Note. Effect sizes (*d*-scores) are calculated for the difference between the current sample and samples from other chapters.

3.1.2 Reliability and Descriptive Statistics for Self-Report EI and Criterion Variables

Internal consistency reliability, descriptive statistics, and gender differences for the SSRI and emotional well-being measures are shown in Table 5.4. SSRI total score and four components were reasonably reliable (components scores for the SSRI were calculated according to the factor solution from Section 3 of this chapter, and scores were computed as the linear sum of items). TAS-20 scores were also reasonably reliable, with externally-oriented thinking showing lower reliabilities than the other two scales, consistent with previous reports (e.g., Barchard & Hakstian, 2004; Chapter 3 of this thesis). Means and standard deviations for the TAS-20 and SSRI cannot be compared to earlier analyses as this study used a 6-point rather than 5-point scale for both measures.

The Satisfaction with Life Scale was not as reliable as reported elsewhere – the fifth item "If I could live my life over, I would change almost nothing" had a negative item total correlation. This item was removed, and further analysis uses a 4-item version of the SWLS. Because McDonald's (1999) division of life satisfaction into current and retrospective includes only two items for retrospective (one of which was item 5), this division was not undertaken in the current study. Loneliness and Stress were both quite reliable measures.

Males scored higher on the externally-oriented thinking facet of the TAS-20, in line with previous research (Hexel, 2003; Montebarocci et al., 2004; also Chapter 3 of this thesis). Females reported greater life satisfaction than males, consistent with research showing females generally report higher levels of well-being than males (e.g., Fujita, Diener, & Sandvik, 1991). Gender differences on the SSRI were confined to the social skills component, where females scored higher than males.

Table 5.4

Reliability, Descriptive Statistics and Gender Differences for Self-Report EI and Criterion Variables

	N	*Alpha*	*Mean*	*SD*	*Gender d*
SSRI	175	.92	147.45	18.14	-0.17
SSRI Factor 1: Optimism/Regulation (11 items)	175	.83	50.15	6.89	-0.10
SSRI Factor 2: Utilisation (5 items)	175	.72	37.99	6.63	-0.02
SSRI Factor 3: Appraisal (9 items)	175	.85	22.85	3.36	0.01
SSRI Factor 4: Social Skills (8 items)	175	.83	36.54	5.81	-0.37*
TAS-20	174				
TAS-20 Difficulty Identifying Emotions (DIF)	174	.79	15.23	5.37	-0.08
TAS-20 Difficulty Describing Emotions (DDF)	174	.78	13.67	5.10	0.11
TAS-20 Externally Oriented Thinking (EOT)	174	.58	13.09	3.73	0.69**
Loneliness	176	.84	23.99	8.79	0.10
Satisfaction with Life	176	.68	21.80	5.44	-0.48**
Satisfaction with Life (items 1 – 4 only)	176	.87	18.51	5.39	-0.42*
Stress	176	.81	19.16	5.07	-0.36

Note. Gender differences are calculated at male – female means, such that negative values indicate higher scores for females.

* $p < .05$, ** $p < .01$

3.1.3 Reliability and Descriptive Statistics for Personality

Table 5.5 shows the descriptive statistics, gender differences, and reliabilities for all 30 facets and five dimensions of the short form of the IPIP version of the NEO-PI-R. Reported reliabilities for the longer version (300 items) and for the actual NEO-PI-R are also shown for purposes of comparison (Costa & McCrae, 1992; International Personality Item Pool, 2001). All scales bar the Dutifulness facet of Conscientiousness have reliabilities over .60, indicating that they can be validly used in research. Reliability estimates for the short form currently used (4 items per facet) are similar to estimates for the longer form (10 items per facet), but clearly lower than for the commercial instrument (i.e., it is a research tool rather than an instrument to be used in industrial-organisational, educational, or clinical domains of psychological practice).

Gender differences in personality are consistent with published results (Costa et al., 2001; Feingold, 1994). Males are higher on the intellect facet of Openness whereas females are higher on Openness to Emotions and Aesthetics. Females score significantly higher on three of the six facets of Agreeableness (Morality, Modesty, and Sympathy) as well as on the broad dimension. Females also score more highly than males on the Neuroticism dimension, with significant differences for Anxiety, Anger, Self-Consciousness, and Vulnerability to Stress at the facet level. In conjunction with the reliability evidence, this suggests that the scale is behaving as expected.

3.1.4 Reliability and Descriptive Statistics for Fluid and Crystallized Intelligence

Table 5.6 shows the reliability, descriptive statistics and gender differences for the Gc and Gf tests as scored conventionally and as scored by consensus (using this sample to determine scoring weights). Consensus scores were created to use in an experimental analysis testing the hypothesis that EI would be distinct from conventionally-scored but not consensus-scored intelligence. Correlations between conventional and consensus scores were high (greater than .75) for all tests except for Nonsense Syllogisms (probably due to the very low reliability of the consensus-scored Nonsense Syllogisms, which may have resulted from the 2-option format).

Table 5.5

Internal Consistency Reliability, Descriptive Statistics, and Gender Differences for the short form of the IPIP-NEO Factors and Facets of Personality (N = 157)

	Descriptive Statistics		*Alpha and Comparisons (IPIP, 2001)*			*Gender d*
	Mean	*SD*	*Current (4 items)*	*IPIP (10 items)*	*NEO (8 items)*	
Openness to Experience	100.95	13.48	.83			-0.16
O1: Imagination	17.54	3.36	.68	.74	.90	-0.11
O2: Artistic Interests	17.63	3.85	.76	.80	.95	-0.43*
O3: Emotion	18.68	3.36	.75	.70	.90	-0.51*
O4: Adventurousness	14.48	3.70	.72	.71	.99	0.07
O5: Intellect	17.84	3.69	.69	.80	.95	0.59**
O6: Liberal	14.78	3.83	.64	.70	.86	-0.08
Conscientiousness	98.08	14.95	.88			-0.16
C1: Self-efficacy	18.04	2.55	.70	.66	.89	0.16
C2: Orderliness	14.01	4.93	.83	.77	.99	-0.06
C3: Dutifulness	18.79	2.61	.50	.60	.87	0.03
C4: Achievement Striving	17.43	3.58	.79	.70	.97	-0.51**
C5: Self-discipline	14.55	3.78	.77	.76	.92	-0.26
C6: Cautiousness	15.25	4.67	.92	.69	.95	0.06
Extraversion	100.06	16.26	.99			-0.02
E1: Friendliness	18.31	3.83	.85	.76	.91	0.07
E2: Gregariousness	16.06	4.37	.80	.78	.98	-0.03
E3: Assertiveness	16.04	3.75	.86	.81	.99	0.32
E4: Activity	15.71	3.30	.62	.70	.98	-0.30
E5: Excitement	15.39	3.97	.78	.67	.95	0.08
E6: Cheerfulness	18.55	3.35	.82	.77	.95	-0.28
Agreeableness	109.41	12.73	.85			-0.71**
A1: Trust	17.66	3.59	.89	.79	.95	-0.01
A2: Morality	19.55	3.08	.73	.64	.86	-0.63**
A3: Altruism	20.31	2.62	.73	.67	.90	-0.36
A4: Co-operation	18.38	3.55	.64	.71	.97	-0.37
A5: Modesty	14.91	3.75	.75	.72	.95	-0.63**
A6: Sympathy	18.60	3.55	.79	.61	.90	-0.62**
Neuroticism	79.10	16.41	.88			-0.63**
N1: Anxiety	14.08	3.99	.78	.75	.90	-0.57**
N2: Anger	12.79	4.41	.87	.76	.91	-0.47*
N3: Depression	10.68	4.14	.84	.80	.92	-0.22
N4: Self-consciousness	13.46	3.93	.66	.72	.94	-0.42*
N5: Immoderation	15.25	3.94	.65	.73	.98	-0.29
N6: Vulnerability	12.85	3.83	.74	.77	.96	-0.66**

Note. Gender differences are male – female mean, such that negative scores indicate female superiority (significance refers to a *t*-test).

*$p < .05$, **$p < .01$

3.1.4.1 Descriptive Statistics for Conventionally-Scored Gf and Gc

The mean for General Knowledge was consistent (2% higher) with previous published reports (Danthiir et al., 2001), as was the mean for Vocabulary (3% higher; Pallier, 2003).[40] All other Gf and Gc tests are short forms of those published elsewhere such that means are not directly comparable (items were selected according to high item-total correlations using data from Kleitman & Stankov, 2005; and MacCann et al., 2004). However, all short forms created for this study have reasonable reliability for experimental purposes. Unfortunately, this is not the case for General Knowledge (used in Kleitman & Stankov, 2005), which has very poor internal consistency that could not be improved by deleting items. Letter Counting was clearly much more difficult than the other five tests – a visual presentation of the letters rather than an auditory presentation probably added to the complexity of the task, and hence increased the difficulty.

3.1.4.2 Descriptive Statistics for Consensus-Scored Gf and Gc

All six cognitive ability tests had lower means and less variability under consensus scoring compared to conventional dichotomous scoring. Reliabilities were similar for all tests except for Nonsense Syllogisms. For Nonsense Syllogisms, consensus scoring resulted in a marked decrease in reliability, probably due to the fact that there were only two options (TRUE or FALSE). Gender differences were the same under both types of scoring: a clear male superiority which contrasted sharply with the female superiority on the STEM. This male superiority on Gf and Gc tests was just as strong when consensus scoring (using this majority-female sample) was used. That is, gender differences in consensus-scored tests were not related to the proportion of each group in the screening sample (a female majority screening sample still resulted in male superiority). Given that there is a significant male superiority for four of the six Gf/Gc markers, and a significant (and quite large) female superiority for one of the EI tests, scores were mean corrected separately for each gender (i.e.,

[40] Note that slightly high means might be expected given that the current sample consists of senior psychology undergraduates in their third year of study and comparison samples consist of first year psychology undergraduates.

male score – male mean and female score – female mean) before further analyses were run, in an attempt to get rid of the possibly confounding effects of gender (see Barchard, 2003 for prior use of this methodology).

Table 5.6

Reliabilities, Descriptive Statistics and Gender Differences for Markers of Fluid and Crystallized Intelligence as Scored Conventionally and by Consensus (Correlation between Conventional and Consensus Scores shown in the far right column)

		Conventional Dichotomous Scores				*Consensus Scores*				*r*
		Descriptive Statistics			*Gender d*	*Descriptive Statistics*			*Gender d*	
		Alpha	*Mean*	*SD*		*Alpha*	*Mean*	*SD*		
Crystallized Intelligence (Gc)										
Vocabulary	(*N* = 164)	.65	.56	.15	0.39*	.69	.50	.08	0.44*	.85
Esoteric Analogies	(*N* = 156)	.70	.72	.16	0.43*	.70	.63	.09	0.44*	.93
General Knowledge	(*N* = 162)	.36	.60	.17	0.16	.34	.52	.08	0.34	.76
Fluid Intelligence (Gf)										
Nonsense Syllogisms	(*N* = 162)	.64	.68	.19	0.50**	.35	.65	.06	0.52*	.57
Letter Counting	(*N* = 110)	.80	.35	.26	0.63*	.77	.16	.09	0.61*	.97
Letter Series	(*N* = 163)	.62	.80	.16	0.09	.64	.60	.09	0.01	.97

Note. Consensus scores for Letter Counting award weights to responses selected by more than 1 person only. Gender differences are male – female mean divided by pooled SD, such that negative scores indicate female superiority (significance refers to a t-test).

*$p < .05$, **$p < .01$

3.2 Structural Models of Intelligence: Relationships between Gf, Gc, and Strategic EI

3.2.1 Correlation Matrix for Gf, Gc and Strategic EI

Correlations between the three EI, three Gc, and three Gf measures are shown in Table 5.7, both for consensus scores, and the usual level scores. Because gender differences were marked, and in the opposite direction for EI than for Gf and Gc, scores were mean corrected separately for males and females before calculating these correlations (although correlations changed very little). Correlations

were similar for consensus and conventional scores, except for the correlation between Syllogisms and Letter Counting, which fell to trivial non-significance when both tests were consensus-scored. The interpretations below refer to the correlations between conventional scores.

Table 5.7

Correlations between EI, Gc and EI Measures when Gc and Gf are Scored Conventionally (bottom right of matrix) and by Consensus (italicised in top left of matrix)

	1	2	3	4	5	6	7	8	9
1. STEU		*.41***	*.38***	*.31**	*.19**	*.33***	*.12*	*.38***	*.15*
2. STEM	.41***		*.39***	*.33***	*.11*	*.31***	*.20**	*.05*	*.01*
3. MEIS	.38***	.39***		*.52***	*.27***	*.54***	*.31***	*.21**	*.09*
4. Vocabulary	.29**	.27***	.47***		*.38***	*.54***	*.27***	*.10*	*.03*
5. General Knowledge	.17**	.07	.24***	.35***		*.46***	*.19**	*.13*	*.01*
6. Analogies	.33***	.30***	.50***	.53***	.39***		*.24***	*.39***	*.26**
7. Syllogisms	.17**	.09	.21**	.35***	.18**	.33***		*.19**	*.01*
8. Letter Series	.36***	.10	.24***	.17**	.19**	.46***	.27***		*.40****
9. Letter Counting	.14	.02	.10	.09	.07	.31***	.20	.40***	

Note. Scores were mean-corrected separately for males and females before calculating correlations. Missing data was dealt with pair-wise, such that *N*s ranged from 96 to 176 (list-wise *N* would be 90).
$^{*}p < .05$, $^{**}p < .01$

Correlations between Gc indicators were consistent with other data sets (Danthiir et al., 2001; Kleitman & Stankov, 2005; MacCann et al., 2004; Pallier et al., 2002). Correlations between the three Gf measures were also consistent with previous findings (Stankov, 2000; Stankov, 2001; Kleitman & Stankov, 2005). The correlations between Gf and Gc indicators confirmed that Esoteric Analogies assesses both Gf and Gc (Kleitman & Stankov, 2005; Roberts & Stankov, 1999). Letter Counting correlations with Gc indicators were low, but consistent with data from MacCann et al. (2004).

EI tests generally correlated more highly with Gc than with Gf, and correlated particularly highly with Esoteric Analogies. Given that Esoteric Analogies is factorially complex, this might indicate that EI test scores also contain elements of both Gf and Gc. Correlations between EI and intelligence are generally stronger for Understanding tests (STEU and MEIS) than for the

Management test (STEM) as has been found in research using the MEIS and MSCEIT (e.g., Mayer et al., 2000b; see Chapter 1 for further references). These differences were only significant for the STEU and STEM's relationship to Letter Counting ($z = 2.34$, $p = .019$), the MEIS and STEM's relationship to vocabulary ($z = 1.97$, $p = .049$), and the MEIS and STEM's relationship to Esoteric Analogies ($z = 2.01$, $p = .044$). That is, there is some evidence that Understanding does relate more strongly to intelligence than Management, even when all tests are in multiple-choice format.

3.2.2 Confirmatory Factor Analysis of Gf, Gc and EI: Three Nested Models

The difference in fit between three nested models – one-factor, two-factor (Gf and EI/Gc) and three-factor (Gf, Gc, EI) – was calculated when Gf and Gc were conventionally scored and again when they were consensus scored. Factor loadings, fit indices, and differences in χ^2 for both these analyses are shown in Table 5.8 (missing data was treated pair-wise).[41] Discrepancy matrices showed an underestimation of the correlation between Esoteric Analogies and Gf measures unless Analogies cross-loaded on Gf, consistent with its status as a factorially complex test. Cross-loadings of Esoteric Analogies onto Gf were thus included in the model.

Only the two-factor and three-factor models in analysis 1 (where Gf and Gc were conventionally scored) had acceptable fit, with RMSEAs not significantly higher than .05 and CFAs above .90 (although still not above .95). For both analyses, χ^2 discrepancies and differences in AIC values indicated that a three-factor model fit significantly better than a two-factor model (which fit significantly better than a one-factor model) suggesting that Gf, Gc, and EI are separate constructs. However, EI and Gc factors were very strongly related, and this relationship was slightly stronger in analysis 2, where all tasks were consensus-scored ($r = .73$ and $r = .79$, respectively).

[41] The analysis was also run with list-wise (N=90) and mean substitution treatment of missing data, and results were similar.

Table 5.8

Factor Loadings, Factor Inter-Correlations, Fit Indices, and χ^2 Differences for One-Factor, Two-Factor, and Three-factor Solutions from Confirmatory Factor Analysis Nested Models of Gf, Gc, and EI

	Analysis 1: Gf and Gc Scored Conventionally						*Analysis 2: Gf and Gc Scored by Consensus*					
	1-factor (27 df)	*2-factor (25 df)*		*3-factor (23 df)*			*1-factor (27 df)*	*2-factor (25 df)*		*3-factor (23 df)*		
Factor Loadings	*g*	*Gc/EI*	*Gf*	*EI*	*Gc*	*Gf*	*g*	*Gc/EI*	*Gf*	*EI*	*Gc*	*Gf*
STEU	.52	.52		.58			.52	.50		.57		
STEM	.43	.48		.54			.47	.49		.53		
MEIS	.64	.69		.73			.70	.71		.75		
Vocabulary	.65	.68			.73		.69	.72			.75	
General Knowledge	.45	.44			.48		.49	.49			.52	
Esoteric Analogies	.79	.50	.41		.54	.42	.78	.64	.31		.65	.36
Letter Series	.48		.67			.69	.36		.75			.77
Nonsense Syllables	.44		.45			.43	.36		.25			.24
Letter Counting	.23		.52			.52	.19		.51			.51
Factor Inter-correlations												
Gf	-	.54	-	.49	.47		-	.36	-	.40	.26	
Gc	-	-	-	.73			-	-	-	.79		
Fit Indices												
RMSEA	.087**	.071		.057			.100**	.090**		.083*		
CFI	.844	.905		.943			.810	.856		.888		
AIC	123.53	108.96		100.52			136.24	124.74		117.606		
χ^2	69.53**	50.958**		38.516*			82.242**	66.743**		55.606**		
$\Delta\chi^2$	-	18.572**		12.442**			-	15.499**		11.137**		

Note. The *p*-value given for the RMSEA is the PLCOSE, a test of the null hypothesis that the RMSEA is no greater than .05.

**$p < .01$, *$p < .05$

Correlations of $r = .70$ are far above Cohen's (1988) definition of "large" and also above the criterion of .70 that Matthews et al. (2002) suggest is "very large" and indicative of substantial overlap of constructs. It certainly seems reasonable to claim that EI (at least in the strategic area) is very strongly related to, if not a part of, Gc. These analyses were run for gender-corrected scores, but an additional set of analyses were run using non-corrected scores to check whether removing gender-related variance substantially changed the model. Results were very similar, although fit indices were better for the gender-corrected model, perhaps indicating that gender-related variation was a source of error. The analysis of non-gender-corrected scores is shown in Appendix 5.2.

3.3 Factor Structure of the Schutte Self Report Scale

In an exploratory factor analysis of the 33 SSRI items, both four-factor and three-factor solutions were extracted, in line with previous research (Petrides & Furnham, 2000 and Saklofske et al., 2003 for four-factor solutions and Austin et al., 2004; Austin et al., 2005, and Gignac et al., 2005, for three factor solutions). Factors were extracted using principal axis factoring and then rotated using the oblimin procedure (a non-orthogonal rotation since the factors are presumably related). The four factor solution accounted for 41.27% of the variance, and the three-factor solution 37.17% of the variance. Factor loadings and communalities from these analyses are shown in Table 5.9.

Table 5.9

Factor Loadings from Exploratory and Confirmatory Factor Analyses of the SSRI Items

Type of Analyses	*Exploratory*									*Confirmatory*							
No. Factors	*3*				*4*					*1*		*3*			*4*		
	1	*2*	*3*	h^2	*1*	*2*	*3*	*4*	h^2		*1/4*	*2*	*3*	*1*	*2*	*3*	*4*
1. When to mention problems	**.48**	.20	.11	.41	.19	.21	.04	**.42**	.44	.64	.64						.70
2. Overcame obstacles	.07	**.43**	.17	.28	.12	.38	.22	-.02	.29	.47		.54			.53		
3. Expect to do well	.01	**.68**	-.06	.45	.01	**.66**	-.02	.03	.45	.50		.62			.62		
4. Easy to confide in	.37	.24	.21	.40	.16	.24	.17	.33	.41	.63	.61						.69
5. Hard to understand non-verbals	**.58**	.08	-.20	.36	**.55**	.09	-.16	.07	.37	.49	.55			.59			
6. Re-evaluate after major events	-.01	.06	**.59**	.37	.03	-.02	**.63**	.03	.41	.32			.58			.58	
7. Mood change = new possibilities	-.09	.07	**.70**	.49	-.17	.01	**.66**	.18	.48	.29			.63			.64	
8. Emotions make life worth living	.32	.00	**.40**	.33	-.11	.00	.26	**.63**	.47	.48	.47						.56
9. Aware of emotions	**.59**	-.11	.24	.41	.13	-.11	.09	**.68**	.55	.55	.59						.67
10. Expect good things	.12	**.42**	.04	.25	-.20	**.47**	-.08	**.44**	.42	.45		.53			.54		
11. Like to share emotions	**.55**	.00	.20	.40	.03	.02	.03	**.75**	.61	.58	.60						.68
12. Make positive emotions last	.27	**.46**	-.05	.39	.17	**.46**	-.05	.17	.40	.58		.61			.61		
13. Arrange events others enjoy	.08	**.47**	.06	.28	.10	**.44**	.10	.02	.28	.47		.54			.54		
14. Seek out happy activities	-.03	.60	.18	.43	-.10	**.57**	.18	.14	.44	.52		.65			.65		
15. Aware of own non-verbals	**.64**	.06	-.13	.42	**.68**	.04	-.05	.01	.48	.56	.61			.69			
16. Create good impression	.16	**.52**	-.02	.36	.14	**.49**	.01	.07	.36	.53		.58			.58		
17. Problems easy in positive mood	.02	**.51**	.21	.36	.16	**.45**	.31	-.12	.40	.51		.57			.57		
18. Recognise facial expressions	**.64**	.13	.05	.53	**.71**	.09	.16	.00	.63	.70	.73			.79			
19. Know why emotions change	**.54**	.05	-.02	.31	.37	.06	-.04	.26	.31	.51	.54			.51			
20. New ideas in positive mood	.04	**.40**	**.40**	.43	.11	.33	**.47**	-.02	.46	.53			.63			.63	
21. Control over emotions	.36	.29	-.23	.29	.34	.29	-.19	.05	.28	.43	.43			.44			
22. Recognise emotions	**.54**	.05	.03	.33	.22	.07	-.07	**.47**	.38	.53	.56						.57
23. Motivate via good outcomes	-.05	**.70**	.06	.48	.01	.66	.12	-.02	.48	.52		.67			.67		
24. Compliment others	.24	.25	.24	.29	.18	.22	.25	.14	.29	.53	.50						.52
25. Aware of others' non-verbals	**.70**	.10	-.13	.53	**.72**	.09	-.05	.03	.59	.64	.69			.76			
26. Empathy	.36	.13	.19	.27	.32	.10	.22	.11	.29	.52	.51			.48			
27. New ideas with emotion change	.16	.10	**.43**	.30	.20	.03	**.49**	.03	.34	.42			.62			.61	
28. Give up because will fail	-.08	**.48**	-.20	.20	-.03	**.47**	-.16	-.05	.19	.21		.33			.33		
29. Read others' body language	**.65**	-.08	.12	.43	**.70**	-.13	.21	.03	.54	.58	.63			.70			
30. Help others feel better	**.50**	.15	.13	.41	.29	.16	.09	.32	.41	.64	.65						.67
31. Good moods as motivators	.06	**.48**	.25	.40	.12	**.42**	.31	-.01	.41	.55		.61			.61		
32. Read others' tone-of-voice	**.66**	-.02	.05	.44	**.63**	-.04	.11	.10	.48	.60	.65			.69			
33. Hard to understand others	**.53**	-.12	-.09	.23	.26	-.08	-.17	.37	.26	.35	.41						.42
Factor Correlations																	
Factor 1												.68	.51		.59	.43	.72
Factor 2													.62			.62	.67
Factor 3																	.52

Note. In the exploratory analysis, loadings > .40 are bolded and loadings > .30 are underlined; Factor 1 = Appraisal of Emotions; Factor 2 = Optimism/Mood Regulation; Factor 3 = Utilisation of Emotions; Factor 4 = Social Skills.

In the four-factor solution, the factors Petrides and Furnham (2000) named "Optimism/Mood Regulation", "Appraisal of Emotions", and "Utilisation of Emotions" were fairly well replicated. However, the "Social Skills" factor was poorly defined, as was the case in Austin et al.'s (2004) analysis. The three-factor solution was essentially the first three of the four factors from Petrides and Furnham (2000) and Saklofske et al. (2003). That is, Social Skills and Emotion Appraisal formed one factor instead of two (and item 8, which loaded on social skills in the four-factor model, loaded on Utilisation of Emotions rather than Appraisal of Emotions in the three-factor model). This agreed with the results Austin et al. (2004) obtained with a 41 item version of the SSRI.

Confirmatory factor analysis (without cross loadings; i.e., each item loaded only on one factor) of one, three and four factor solutions were also conducted to test whether positing more factors significantly improved the fit of the model. Factor loadings and correlations from these analyses are also shown in Table 5.9. The four-factor model fit significantly better than the three-factor model ($\Delta\chi^2(3) = 110.434$, $p < .001$), which fit significantly better than the one-factor model ($\Delta\chi^2(3) = 210.573$, $p < .001$). All loadings in all three models were significant at $p < .01$. However, the fit of the four-factor model was still not good (GFI = .729, CFI = .742; RMSEA = .083, $p_{close} < .001$).[42] In further analysis, four SSRI component scores derived from the linear sum of the items loading most highly on each of the four factors were used, since the four-factor solution showed the best fit.

3.4 Assessing the Relationship of Performance-Based and Trait EI to Personality

3.4.1 Trait EI and Personality

Table 5.10 shows the correlations of NEO-PI-R facet scores with the four SSRI factors and with scores on four performance-based EI tests. For reasons of space, correlations among the 30 personality facets are not shown here but are available in Appendix 5.3. The Appraisal subscale of the

[42] Since gender differences were observed for the SSRI total score, analyses were also conducted controlling for gender by mean correcting each item separately for males and females. Fit indices and differences between models were similar for these gender-controlled models.

SSRI showed no substantial relationships with any facet of personality (largest correlation is $r = .33$), indicating its distinctiveness as a construct. Utilisation of Emotions was also reasonably independent from personality, although there is a clearer pattern of relationships: Utilisation relates to the non-intellectual facets of Openness (Imagination, Artistic Interests, and Openness to Emotion) and most facets of Extraversion. Optimism and Social Skills were substantially replications of Extraversion (with some other strong relationships to facets from other personality traits), in agreement with Saklofske et al. (2003). The Social Skills SSRI factor related most strongly to Openness to Emotions, and also related reasonably strongly to Friendliness and Altruism.

Essentially, it appears that two of the four factors of the SSRI are not independent of personality. The Mood Regulation/Optimism factor is an Extraversion/Neuroticism composite, and the Social Skills factor is an Extraversion/Openness composite, with strong relationships to other narrow conceptualisations of personality. However, the Appraisal and Utilisation factors of the SSRI are relatively independent of personality. It might thus be these factors that stand the best chance of providing incremental prediction of emotional well-being beyond personality.

Table 5.10

Correlations of Personality Factors with SSRI Factors and Performance-based EI Scores

	Trait EI: SSRI					*Performance-based EI*		
	Fac1: App	*Fac2: Opt*	*Fac3: Use*	*Fac4: Skills*	*Total*	*STEU*	*STEM*	*MEIS*
Openness to Experience	.18*	.22*	.29**	.23**	.28**	.21*	.24**	.27**
O1: Imagination	.05	.16	.23**	.04	.14	.12	.29**	.17*
O2: Artistic Interests	.26**	.26**	.35**	.22**	.33**	.07	.18*	.18*
O3: Emotion	.27**	.20*	.33**	.54**	.42**	.10	.16	.02
O4: Adventurousness	.10	.23**	.11	.10	.18*	.08	.02	-.01
O5: Intellect	.11	.21*	.14	.02	.15	.27**	.14	.36**
O6: Liberalism	-.11	-.23**	-.08	-.04	-.16	.13	.10	.26**
Conscientiousness	.17*	.28**	-.08	.23**	.23**	.00	.01	-.18*
C1: Self-efficacy	.19*	.51**	.09	.25**	.36**	.16	.14	.03
C2: Orderliness	.09	.03	-.04	.06	.06	-.18*	-.18*	-.24**
C3: Dutifulness	-.03	.14	-.06	.11	.07	.08	.02	-.12
C4: Achievement Striving	.17*	.35**	.07	.37**	.33**	.04	.07	-.10
C5: Self-discipline	.20*	.37**	.01	.26**	.30**	-.02	.02	-.18*
C6: Cautiousness	.06	-.08	-.28**	-.05	-.08	.06	.05	-.04
Extraversion	.26**	.66**	.35**	.37**	.54**	.05	.12	-.05
E1: Friendliness	.33**	.52**	.34**	.45**	.54**	.01	.13	-.11
E2: Gregariousness	.28**	.48**	.33**	.30**	.45**	.02	.08	-.04
E3: Assertiveness	.20*	.56**	.16	.21*	.38**	.05	.02	.00
E4: Activity	.01	.40**	.21*	.18*	.25**	.10	.09	-.04
E5: Excitement	.08	.34**	.24**	.16	.26**	.02	.02	.00
E6: Cheerfulness	.20*	.55**	.23**	.32**	.43**	.00	.22**	-.06
Agreeableness	.13	.07	.16	.28**	.20*	.11	.13	.02
A1: Trust	.13	.19*	.06	.21*	.20*	.07	.02	-.02
A2: Morality	-.02	.13	.14	.15	.12	.20*	.20*	.02
A3: Altruism	.30**	.25**	.31**	.45**	.41**	.08	.10	.06
A4: Co-operation	.12	.06	-.02	.12	.10	.13	.13	-.04
A5: Modesty	-.19*	-.44**	-.04	-.15	-.30**	-.04	-.02	.04
A6: Sympathy	.23**	.18*	.23**	.39**	.33**	.04	.11	.04
Neuroticism	-.32**	-.44**	.00	-.09	-.31**	-.09	-.03	.08
N1: Anxiety	-.12	-.32**	.02	.07	-.13	-.15	-.04	-.03
N2: Anger	-.32**	-.07	.02	.00	-.14	-.20*	-.12	.04
N3: Depression	-.20*	-.51**	.04	-.12	-.30**	.04	-.05	.09
N4: Self-consciousness	-.29**	-.53**	-.24**	-.32**	-.46**	.02	-.04	.06
N5: Immoderation	-.15	.06	.18*	.00	.01	.05	.15	.13
N6: Vulnerability	-.19*	-.43**	-.02	.03	-.22**	-.12	.00	.04

** $p < .01$, * $p < .05$

3.4.2 Performance-based Strategic EI and Personality

As hypothesised, correlations between Strategic EI and personality traits and facets were all less than $r = .30$ (except for the MEIS composite and Intellect, $r = .36$). In contrast to Chapter 3 (where only Agreeableness related to the STEU and STEM), only Openness to Experience correlated with Strategic EI scores in this chapter, and most correlations at the facet level were with Openness. In contrast to hypotheses, no Strategic EI measure correlated significantly with Openness to Emotions. In fact, the strongest correlations with personality for both the STEU and the MEIS were with Intellect, whereas STEM scores correlated most strongly with Imagination. If Openness is considered separately as an intellectual component (i.e., Intellect, Liberalism and Adventurousness facets) and an emotional component (i.e., Emotion, Artistic Interests and Imagination facets; see Gignac et al., 2004 for empirical evidence for this division), relationships between EI and Openness become clearer. Management does not relate to the intellectual component, but only to the emotional component (although not directly to the Emotions facet). However, Understanding relates more strongly to the intellectual than the emotional component (although the MEIS composite does show some smaller correlations with facets from the emotional component). For Understanding, there is no support for the hypothesis that EI would relate more strongly to emotion-related facets of Openness than to intellectual facets. For Management, there is some support for this hypothesis, although relationships were found for the Imagination rather than Emotions facet of Openness.

Strategic EI scores related to only one of the six facets of Agreeableness. Both STEU and STEM showed a small degree of relationships to Morality. Given that both Morality and STEM scores were significantly higher for females ($d = -.76$ and -.71, respectively), it is possible that gender differences on the STEM were may be due to the shared variation with Morality. An ANCOVA controlling for Morality still found significant gender differences on the STEM ($F_{1,138} = 9.598$, $p = .002$), with marginal means estimated as $M = .62$ ($SE = .01$) for men and $M = .67$ ($SE = .01$) for women.

Relationships of the STEU and STEM to the facets of Neuroticism were mostly trivially small and non-significant (only the STEU showed any significant correlation and this was small, $r = .20$ with Anger). This result indicates that the relationships with state depression, anxiety, and stress found for the STEU and STEM in Chapter 4 are not due to the STEU and STEM's overlap with trait Depression, Anxiety, and Vulnerability.

3.5 Comparing the Incremental Prediction of Performance-based and Trait EI

Two regression analyses were carried out to predict each criterion variable: (1) a simple regression analysis predicting from EI alone; and (2) a hierarchical regression analysis accounting for intelligence and personality as well as EI. These two regression models were undertaken separately for each criterion variable (loneliness, life satisfaction, stress, and the three components of the TAS-20 scale of alexithymia), and for each EI predictor (STEU, STEM, MEIS and the four SSRI scores). In the hierarchical regression, Gf and Gc composite scores were entered as the first step.[43] In the second step, the five-factor model dimensions of personality were added. In the final step, the EI measure of interest (STEU, STEM, MEIS or SSRI score) was added to the model. R^2 estimates are shown for the simple regression and for each step of the multiple regression in Tables 5.11 and 5.12. Standardised beta weights are given in Appendix 5.4 (rather than in the tables) for reasons of space. These analyses test whether each measure of EI predicts criteria at all, and whether scores incrementally predict criteria over and above the prediction by personality and intelligence.

3.5.1 Prediction of Alexithymia from Performance-based and Trait EI

Table 5.11 shows the squared multiple correlation (R^2) after each step when predicting the three components of alexithymia: Difficulty Identifying Feelings (DIF), Difficulty Describing Feelings (DDF) and Externally Oriented Thinking (EOT). Since R^2 values are the same until the point where the EI score in entered, values for steps 1 and 2 are presented only once, and the differing

[43] Gf and Gc factor scores from a two-factor exploratory factor analysis were used as Gf and Gc, as using 6 different indicators of cognitive ability might be a problem for collinearity.

values for step 3 are given separately for each EI variable (STEU, STEM, MEIS composite and the four SSRI component scores). The R^2 for EI's prediction of criteria without controlling for intelligence or personality is also shown in the last line of Table 5.11.

Table 5.11

Squared Multiple Correlations (R^2s) and Incremental R^2s for Three-Step Hierarchical Regression Models Predicting Alexithymia from Intelligence in Step 1, Personality in Step 2, and EI in Step 3(R^2 for Simple Linear Regressions Using EI Only Also Shown)

	Step 1 (Gf + Gc)	*Step 2 (FFM)*	*Step 3 (EI)*						
			STEU	*STEM*	*MEIS*	*AES1*	*AES2*	*AES3*	*AES4*
Regressions Predicting Difficulty Identifying Feelings (DIF)									
R^2	.03	.47**	.47**	.48**	.49**	.51**	.50**	.49**	.56**
ΔR^2	.03	.44**	.00	.01	.02	.04*	.03	.02	.09**
R^2 *(EI only)*[a]			.00	.00	.00	.05**	.08**	.00	.12**
Regressions Predicting Difficulty Describing Feelings (DDF)									
R^2	.08	.31**	.32**	.31**	.31**	.35**	.35**	.32**	.58**
ΔR^2	.08	.23**	.02	.00	.01	.04*	.04*	.01	.27**
R^2 *(EI only)*[a]			.00	.00	.00	.08**	.06**	.02	.34**
Regressions Predicting Externally Oriented Thinking (EOT)									
R^2	.02	.45**	.46**	.47**	.48**	.48**	.49**	.47**	.48**
ΔR^2	.02	.43**	.01	.01	.03	.02	.04*	.01	.03
R^2 *(EI only)*[a]			.01	.07**	.02	.13**	.13**	.17**	.33**

Note. For the multiple regression, missing data was deleted list-wise (N = 74). AES1 = Appraisal of Emotions; AES2 = Optimism/Mood Regulation; AES3 = Utilisation of Emotions; AES4 = Social Skills.

[a]R^2 *(EI only)* refers to the squared zero-order correlation between the criterion variables and EI (i.e., intelligence and personality are not accounted for); for this set of analyses, N = 170.

*$p < .05$, **$p < .01$.

Of the three Strategic EI measures, only the STEM predicted any of the components of alexithymia, either alone or after intelligence and personality were controlled for. The STEM showed a small relationship with EOT, in agreement with results from Chapter 3. With the exception of Utilisation, all SSRI components predicted all three components of alexithymia before controlling for intelligence or personality (Utilisation only predicted EOT). After controlling for personality, only Optimism significantly predicted EOT, Appraisal and Social Skills both predicted DIF and Utilisation, Appraisal and Social Skills all still predicted DDF. However, the only case where greater than 5% of the variance was explained was for Social Skills (which uniquely accounted for 9% of the variance in DIF and 27% of the variance in DDF). The 27% of shared variance between Social Skills and DDF implies that these are quite possibly assessing the same construct (the semi-partial correlation would be $r = .52$, which is above Cohen's [1988] criteria for a large effect size).

Essentially, the SSRI predicts TAS-20 scores much more strongly then the measures of performance-based EI do. None of the STEU, STEM or MEIS provided any incremental prediction of alexithymia components. Although the SSRI did relate to the TAS-20 scores even after controlling for personality and intelligence, one of the components of the SSRI related so strongly to DDF that it appears to be measuring the same construct.

3.5.2 Prediction of Loneliness, Life Satisfaction, and Stress by Performance-based Strategic EI

Table 5.12 shows the same regression models run to predict loneliness, life satisfaction and stress. Although the SSRI components predicted all three criteria reasonably strongly in simple regressions, all but one of these relationships reduced to non-significance after personality was accounted for. Appraisal predicted an additional 3% of the variation in Loneliness. None of the Strategic EI scores predict loneliness, life satisfaction, or stress at all, even before controlling for intelligence or personality. This lack of prediction of life satisfaction and stress contrasts with Chapters 3 and 4, where STEM scores predicted both constructs. However, the SWLS was given as a state measure in the current study (with participants instructed to answer as they felt "right now at this moment") whereas no such time restrictions were placed on responses in Chapter 3. This discrepancy

may also be due to the use of short forms of the STEU and STEM and the correspondingly lower reliability estimates in the current study. However, intelligence factors (Gf and Gc) also did not significantly predict life satisfaction, loneliness or stress, which were strongly predicted only by personality factors. This points to a possible measurement method shared by predictors and criteria: The three criteria are all self-report rating scales, as are the personality indicators.

Table 5.12

Squared Multiple Correlations (R^2s) and Incremental R^2s for Three-Step Hierarchical Regression Models Predicting Loneliness, Life Satisfaction and Stress from Intelligence in Step 1, Personality in Step 2, and EI in Step 3(R^2 for Simple Linear Regressions Using EI Only Also Shown)

	Step 1 (Gf + Gc)	*Step 2 (FFM)*	*Step 3 (EI)*						
			STEU	*STEM*	*MEIS*	*AES1*	*AES2*	*AES3*	*AES4*
Regressions Predicting Loneliness									
R^2	.03	.47**	.48**	.49**	.47**	.50**	.49**	.47**	.49**
ΔR^2	.03	.44**	.01	.02	.00	.03*	.02	.00	.03
R^2 *(EI only)*[a]			.01	.02	.01	.10**	.32**	.03*	.18**
Regressions Predicting Life Satisfaction									
R^2	.01	.21*	.21*	.23*	.22*	.21*	.22*	.21**	.21*
ΔR^2	.01	.19*	.00	.03	.02	.00	.01	.00	.01
R^2 *(EI only)*[a]			.00	.01	.00	.01	.12**	.00	.06**
Regressions Predicting Stress									
R^2	.11**	.31**	.31**	.31**	.32**	.31**	.31**	.31**	.31**
ΔR^2	.11**	.20**	.00	.01	.01	.01	.00	.00	.00
R^2 *(EI only)*[a]			.02	.00	.01	.03*	.05**	.01	.00

Note. For the multiple regression, missing data was deleted list-wise (N = 74). AES1 = Appraisal of Emotions; AES2 = Optimism/Mood Regulation; AES3 = Utilisation of Emotions; AES4 = Social Skills.

[a]*R^2 (EI only)* refers to the squared zero-order correlation between the criterion variables and EI (i.e., intelligence and personality are not accounted for); for this set of analyses, N = 170.

*$p < .05$, **$p < .01$.

4. Discussion

Psychometric evaluation of the STEU and STEM yielded the hypothesised results: Test-retest reliability was higher than internal consistency, and females scored more highly on the STEM. Hypotheses regarding a strong relationship between Strategic EI and Gc were also supported, with a very strong relationship between Gc and EI factors. However, a four-factor structure for the SSRI fit better than a three-factor structure, going against hypotheses. Performance-based EI measures did not incrementally predict any of the emotional well-being measures and were only weakly related to personality, whereas SSRI components incrementally predicted some criteria, and two of the four components showed substantial overlap with personality. These issues are discussed in further detail below.

4.1 Reliability of Performance-Based EI: Internal Consistency versus Test-Retest Reliability

Reliabilities of the three performance-based measures of Strategic EI were low. Given that these were short forms of the measures used in Chapter 3 this may not indicate a huge problem for test development. In addition, test-retest reliability was higher than internal consistency reliability (in agreement in Hypothesis 1), which might indicate that the tests assess several dimensions. Factorially-complex tests such as measures of general intelligence are known to produce poor internal consistency estimates (e.g., Gregory, 2004), such that stability over time is greater than stability over different sets of items. In the current study, test-retest correlations were higher than internal consistency even though conditions were fairly unfavourable for test-retest reliability. Students completed the EI tests as part of a class project on EI, and learnt about EI theories and the STEU and STEM's development in the intervening weeks between administrations of the STEU and STEM. Thus background factors (such as how much students had attended class and paid attention to the material) may have differentially affected their scores on the second administration of the STEU and STEM, reducing the similarity of scores between the first and second administrations.

The SJT method used to develop the STEM may have resulted in a multi-dimensional test. SJTs often assess quite complex knowledge of types of behaviours and as such may be multi-

dimensional, such that internal consistency is low even though tests are stable over time (Chan & Schmitt, 2002). Another possible reason for multi-dimensionality is the multiple emotions covered by both the STEU and STEM. Content in the STEU and STEM covers several emotions (14 in the STEU and 3 in the STEM), with both MEIS tests used also dealing with several emotions. Some commentators have recently suggested that processing of emotional information may be different for different emotions (e.g., O'Sullivan & Ekman, 2004) and newer instruments have incorporated this idea into test development. For example, the Personal Introspection of Emotional States (PIES) self-report scale includes separate scales for positive and negative emotions (Roberts et al., 2005). The idea that components of EI may be specific to different types of emotions is examined in Chapter 6, which briefly compares emotion-specific versus general structures for the STEU and STEM.

The reliability of the STEU and STEM found in this and previous chapters highlight the importance of testing reliability over multiple different settings: over time, over groups of participants, and over subsets of items. Results from Chapter 4 indicate that STEU and STEM may be less generalisable to a highly educated adult sample, although this could also be due to restriction of range issues. Results from this sample suggest that tests are more stable over time than they are internally consistent, raising important questions about whether abilities involved with processing emotions *can* be uni-dimensional or may involve different processes for different emotions.

4.2 Group Differences on the STEU and STEM

Different cultural groups *did* show different scores on the STEU and STEM, with the majority culture (Australian/Anglo-Celtic) performing better than other cultural groups. However, controlling for English speaking experience eliminated these differences, indicating that the advantage for the majority group was due to superiority at the English language rather than exposure to a particular set of cultural norms. In conjunction with the high level of relationship between Strategic EI and Gc, this finding might indicate that at least some of the variation in Strategic EI scores is due to language or verbal ability. An important question is *how much* of the variation in STEU and STEM scores is due to this language-ability-variance, rather than variation due to EI. In the following chapter (Chapter 6)

this issue is examined in more detail, by controlling for reading comprehension ability when looking at the relationship between EI measures, and between EI, Gf, and Gc.

Women scored more highly on the STEM, as hypothesised. One possible reason for this female advantage might be the verbal load required to interpret the STEM's items. Women have higher verbal ability than men and may score more highly on the STEM due to their greater ability to comprehend the items (e.g., Halpern & LaMay, 2000). However, males in this sample scored significantly more highly on measures involving verbal ability (Vocabulary and Esoteric Analogies tests) such that higher STEM scores for females cannot be due to higher verbal ability in this case.

Another potential reason for female superiority on tests which involve the processing of emotional information is the female superiority in memory tasks that are episodic or autobiographical (i.e., memory of "real life" situations; Herlitz, Nilsson, & Backman, 1997; Seidlitz & Diener, 1998). Compared to men, women accurately recall more of their own emotional life events, in more vivid detail, and are faster at accessing these memories (Davis, 1999; Ross & Holmberg, 1992; Seidlitz & Diener, 1998). Davis found that this difference obtained only for *emotional* memories. A consequence of this is that women may be more able to call up analogues from their own lives to help answer items on tests such as the STEM. With better memory of emotional situations, women may also be working from a wider knowledge base when it comes to emotional situations. The fact that gender differences are stronger for Management than Understanding, and that Management test items more strongly resemble autobiographical or episodic content supports this idea.

Although Gc tasks are generally constructed such that there are no gender differences, general knowledge tasks (which may not constructed to be gender neutral) show a marked male superiority in most domains (Ackerman, Bowen, Beier, & Kanfer, 2001; Lynn, Irwing, & Cammock, 2001). The female superiority at knowledge of how to manage emotional situations is thus an important difference from other knowledge-based tasks. It is a particularly important consideration for applied uses of psychological tests where adverse impact may be an important consideration. It is commonly argued that EI must show incremental prediction over existing constructs such as personality and

intelligence to be of use in applied settings (e.g., Brody, 2004; Izard, 2001; Wilhelm, 2004). However, EI test scores may be useful for selecting groups with less adverse impact (at least in terms of gender) rather than in their incremental prediction. As a rival predictor to other measures of cognitive ability, EI's strength may lie in selecting equally effectively, but from different groups of people, rather than adding incremental prediction. Naturally this argument rests on the assumption that EI should predict criteria just as well as existing measures of intelligence do. However, this is a particularly important argument given that both general knowledge inventories and Raven's Progressive Matrices show a clear male superiority and are very frequently used in applied settings as measures of intelligence (Colom, Escorial, & Rebollo, 2004; Lynn et al., 2001; Lynn & Irwing, 2004).

4.3 Relationships between Strategic EI and Fluid and Crystallized Intelligence

4.3.1 Emotion Understanding is More Cognitively Saturated than Emotion Management

Using multiple-choice forms of both the STEU and STEM meant that differences between Understanding and Management's correlations with intelligence could not be due to response format (which is possible in the MSCEIT where Understanding and Management tests are in different formats). Current results confirmed that Understanding relates more strongly to intelligence than Management does, even when both Understanding and Management are measured with multiple-choice indicators. The MEIS and STEU related more strongly to both Gf and Gc than did the STEM, in agreement with analyses of MSCEIT Understanding and Management (e.g., Bastian et al., 2005). However, the difference in Management-Gc and Understanding-Gc correlations in this study was smaller than has been reported for the MEIS and MSCEIT (Bastian et al., 2005; Lam & Kirby, 2002; Lopes et al., 2003). The STEM related to intelligence (particularly Gc) more strongly than MSCEIT Management tests have done, implying that the relationship of EI tests to intelligence *is* affected by response format (with stronger relationships for multiple-choice tests than ratings-based tests). Chapter 3's more direct comparison of response formats for the STEM support this conclusion. Essentially, current findings support previous reports of greater relationships of emotional

Understanding to intelligence (compared to the other branches). However, results also suggest that the magnitude of this difference may have been exaggerated by response format differences between Understanding and Management tests.

4.3.2 Distinction of Strategic Performance-Based EI from Gf and Gc

All three components of Hypothesis 3 were supported: (1) EI formed a distinct factor from Gf and Gc; (2) this EI factor was more strongly related to Gc than to Gf; and (3) EI and Gc factors were more strongly related when Gc tests were also consensus scored. However, EI and Gc factors were very strongly related, and the difference between analyses of consensus scores and conventional scores was slight.

Davies et al.'s (1998) suggestions that consensus-scoring may act as a method effect in distinguishing EI from intelligence were supported by current results: Consensus-scored EI related more strongly to Gc when Gc tests were also consensus scored than when Gc tests were scored conventionally. However, the strength of correlation between EI and Gc factors increased only slightly when Gc tests were scored by consensus (from .73 to .79). That is, there is evidence that consensus scoring is partly responsible for the factorial distinctiveness of EI but the effect size is small.

Given that EI tasks defined a factor distinct from but highly correlated with Gc, EI's distinction from Gc probably cannot be decided from structural evidence alone. Horn and Noll (1997) suggest several additional types of evidence that distinguish between different kinds of cognitive abilities: (1) developmental-functional differences and age trends; (2) neuro-cognitive differences (i.e., different relationships to physiological and neural functioning); (3) prediction of different types of achievements with different strengths; and (4) heritability differences. They include different personality correlates and group differences in their definition of developmental-functional distinctions, but these could easily be considered two additional parameters in their own right. Types of evidence of EI's status as a distinct intelligence examined in this thesis include structural, developmental, achievement-related, personality correlates, and group differences criteria.

If EI is a part of Gc then age trends should echo Gc, with a small steady increase during the 20s and 30s (see e.g., Schaie, 1994). This increase from young to middle adulthood appears to be the case from the data sampled so far. The community sample of median age 33 scored significantly more highly on both STEU and STEM than the first year university sample of median age 19. However, sampling issues in selection of the community sample (e.g., high level of education, motivation and computer literacy) mean that these results are in need of replication. In addition, it is not possible with the current data to test for the large increases in intelligence expected during the primary and secondary schooling years (age about 7 to 17), which would provide further support for this proposition. In the current chapter, correlations with personality are the same for EI as prior research has found for Gc, both relating most strongly to the Openness (Ackerman & Heggestad, 1997). However, both these similarities were stronger for Understanding than Management measures. The community sample scored relatively more highly on the STEU than the STEM, and both STEU and MEIS scores related most strongly to the Intellect facet of Openness, whereas the STEM correlated most highly with the Imagination facet (none of the measures correlated significantly with Openness to Emotions, contrary to hypotheses). In addition, Chapter 3 found that the STEM correlated reasonably highly with Agreeableness and trivially negatively with Openness (r = .23 and r = -.11 respectively), whereas the STEU's correlations were trivially positive for both Agreeableness and Openness (r = .16 and r = .10 respectively). Both age and correlate trends indicate that Understanding is *not* distinct from Gc. The case is not so strong for Management, particularly given that females score more highly on Management tests and males on most other knowledge tests. Generally, results support the idea that Strategic EI is probably part of crystallized intelligence. As such, claims that EI differs in important ways from other content-bound abilities (e.g., Gv or Ga) may not be correct, at least for the Strategic area of EI.

Some of this overlap between EI and Gc may be due to the text-based format of Strategic EI items. Getting to the correct answer requires not only the appropriate emotion-related knowledge, but the ability to comprehend the text of the item and options. The fact that cultural differences

disappeared once English language experience was accounted for supports the idea that some of the variation in scores is simply due to the ability to comprehend the item's text.

Confounding of socio-emotional knowledge with the text-based format used to assess such knowledge was a major (and unresolved) problem for previous research into social intelligence (SI, see Kihlstrom & Cantor, 2000, for a review). Although the issues are similar for EI as they were for SI, technology has become increasingly sophisticated since much of the earlier SI test construction after Guilford's model (e.g., Guilford, Hendricks, & Hoepner, 1968; O'Sullivan & Guilford, 1975; O'Sullivan & Guilford, 1976). The possibility of presenting scenarios as high quality film footage, animations, or in auditory format (rather than text) might allow testing of whether the overlap of EI with intelligence is intrinsic to the EI construct, or are based on the verbal format that items are presented in. Alternatively, studies can be designed to assess the impact of text-reading on what text-based EI tests measure. Chapter 6 examines the relationship of text-based and non-text-based measures of EI, and also the relationship of the STEU and STEM (text-based) to verbal and nonverbal measures of intelligence, after controlling for reading comprehension ability.

4.4 Factor Structure of the SSRI

Hypothesis 4 (that a three factor rather than four-factor solution would best describe the SSRI) was not confirmed. Although three factors equivalent to previous analyses could be extracted (Austin et al., 2004; 2005; Gignac et al., 2005), a four-factor solution fit better than a three-factor solution. In any case, Petrides and Furnham's (2000) assertion that the SSRI was not uni-dimensional is indubitably correct. One of the four factors (Optimism/Mood Regulation) is not independent of Extraversion, correlating highly with all Extraversion facets. As a third of the SSRI's items define Optimism/Mood Regulation and this component correlates $r = .83$ with the SSRI total score, any analyses using the SSRI as one dimension is substantially contaminated by Extraversion.

Austin et al.'s (2004) extension of the SSRI to include more negatively keyed items and more items representative of each of the three components could be further extended to include more items in the Appraisal, Utilisation, and Social Skills components, to the exclusion of Optimism-related

items (as these are both conceptually and empirically equivalent to Extraversion). With more items defining each component, an examination of whether trait Optimism can be distinguished from trait Mood Regulation would be possible. Further specification of what the Social Skills component is actually measuring would also be useful, as this component was uncovered empirically and was not part of the original definition that Schutte et al. (1998) used to develop SSRI items. This is a particularly important issue given that Social Skills correlated so highly with the Difficulty Describing Feelings.

4.5 Incremental Validity of Trait and Performance-Based EI

The prediction of emotional well-being variables was unimpressive for the performance-based measures, possibly due to poor reliabilities. Only the STEM related to externally-oriented thinking (which both STEU and STEM were related to in Chapter 3) and this was not significant when personality and intelligence were controlled. There were no significant relationships between Strategic EI and life satisfaction, loneliness, and stress, or to the other two components of alexithymia.

The four SSRI components did slightly better than the performance-based EI measures in predicting emotional well-being, particularly aspects of alexithymia. Two components incrementally predicted Difficulty Identifying Feelings, three components incrementally predicted Difficulty Describing Feelings and one component incrementally predicted Externally Oriented Thinking. However, there was no incremental prediction of Life Satisfaction or Stress and only one of the components (Appraisal) predicted Loneliness, explaining a modest additional 3% of variation. One reason that self-report measures of EI may be superior in predicting self-report measures of emotional well-being is that they *are* actually measures of emotional well-being – the predictor and criterion are not sufficiently distinct (e.g., Roberts et al., 2004; Wilhelm, 2004). This might particularly be the case for the Social Skills component of the SSRI. Although the first three factors of the SSRI were constructed according to theory, Social Skills came out of empirical analysis, and some items are semantically very similar to emotional well-being measures. For example items “I like to share my

emotions with others" and "I easily recognize my emotions as I experience them" are from the SSRI's Social Skills whereas "It is difficult for me to reveal my innermost feelings, even to close friends" and ". It is difficult for me to find the right words for my feelings" are from the Difficulty Describing Feelings component of the TAS-20. These results illustrate the point that criterion measures should at the very least be distinguishable from the variables used to predict them. Observational data and more objective measures (e.g., cortisol output or heart rate changes under stressful situations) might be used in future studies as an adjunct or replacement for rating scales.

4.6 General Conclusions

It would appear that Strategic EI is an alternative content area (i.e., knowledge of emotional phenomenon) for the measurement of crystallized intelligence, whose possible utility lies in the group differences rather than the incremental prediction. Strategic EI and in particular Emotion Management appear to be an area where women score more highly than men. If EI predicts similar outcomes to traditional measures of intelligence, the use of EI tests may reduce adverse impact on women, or at least ameliorate the male advantage in several cognitive tests currently used. Trait EI appears to be an expansion and re-categorisation of the existing emotion-related personality facets. Emotion appraisal (i.e., the perception/recognition of emotional content) was the component of the SSRI most clearly distinct from personality, echoing the findings of Davies et al. (1998) in their examination of performance-based EI and intelligence. It seems that the unexamined territory of EI (the parts of EI unrelated to personality and intelligence) may lie in the emotion perception area for both trait and performance-based EI. This issue is explored in the following chapter, which includes performance-based measures of facial and vocal expression recognition as possible indicators of a general EI factor.

Chapter 6

Do Strategic EI and Emotion Recognition Ability Form a Unified "Emotional Intelligence" Construct?

1. Introduction to Chapter 6

So far, this thesis has examined Strategic EI as measured by text-based tests (the STEM, STEU, and component tests from the MEIS battery). However, there is growing use of emotion recognition ability (ERA) tests as alternative ways to conceptualise and measure EI (Austin, 2004; Roberts et al., submitted). Determining whether ERA and text-based EI measures measure the same construct, and whether this construct can be subsumed under existing theories of intelligence are the two primary aims of this chapter. Secondary aims include examining possible emotion specificity in Understanding and Management constructs; checking how much of the variation in STEU and STEM scores is due to reading comprehension skill; and re-examining EI's relationship to personality and to academic achievement. These issues are outlined in further detail below.

1.1 Examining Emotion Specificity in the STEU and STEM

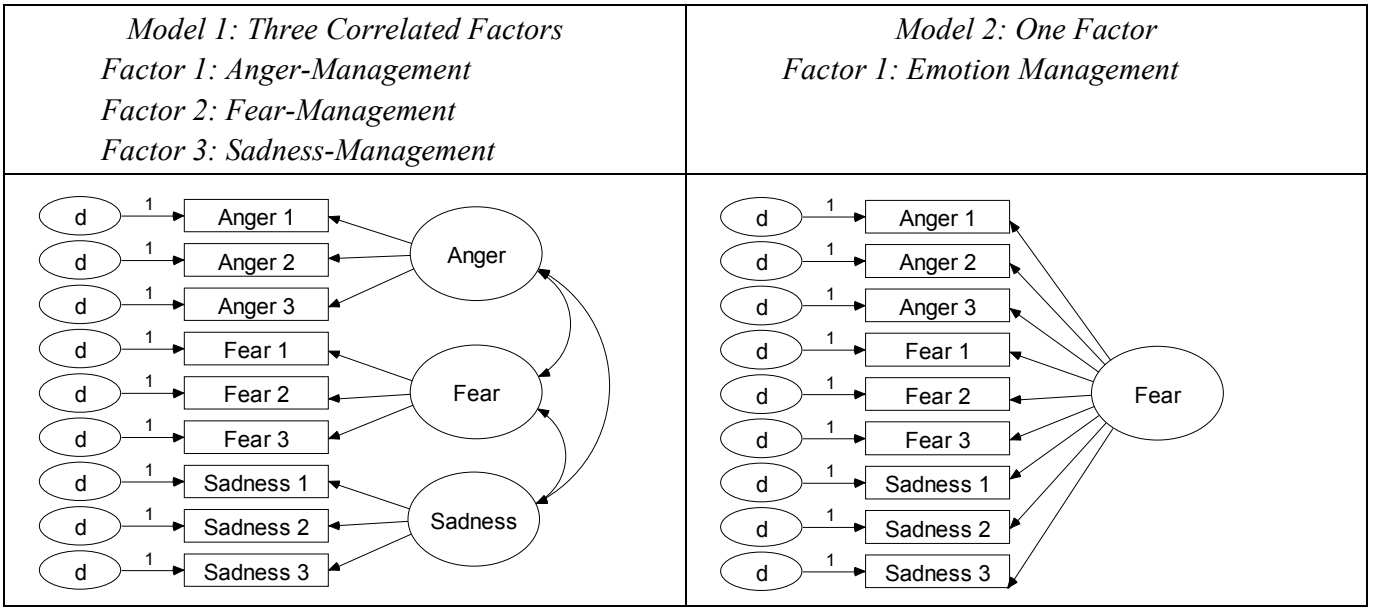
Results from Chapter 5 indicated that the test-retest reliability of both the STEU and the STEM was higher than the internal consistency, suggesting that tests might be multi-dimensional. O'Sullivan and Ekman (2004) recently suggested that the processing involved in emotional tasks might be different for different emotions. If this is the case for Understanding and Management then this may explain why the STEU and STEM are multi-dimensional, since the STEU involves fourteen different emotions and the STEM three. In the current study, this possible source of multi-dimensionality will be examined empirically, by ascertaining whether item parcels representing different emotions load on distinct emotion-specific factors or on one general factor. Since the fourteen emotions covered by the STEU would represent too complex a model to examine empirically, additional STEU items representing sadness, anger, and fear were created such that

three-factor emotion specific models (i.e., fear, anger and sadness) could be examined for both the STEU and the STEM.

The design for this analysis is outlined in Figure 6.1. For both the STEU and the STEM, nine item parcels were used as indicators in a series of structural analyses. Three indicators represent anger content (i.e., understanding anger for the STEU or anger-management for the STEM), three represent fear content, and three represent sadness content. In Model 2, anger, fear, and sadness indicators form three separate factors. Model 2 is nested within Model 1, where all nine indicators load on the same factor, such that the increase in model fit for positing emotion specificity can be statistically evaluated. Based on evidence of multi-dimensionality in Chapter 5, and on suggestions of emotion specificity in the EI literature, it is expected that emotion specific models of Understanding and Management will provide a better fit to the data than one-factor general models.

Figure 6.1

Structural Models Testing Emotion Specificity Conducted for STEM Item Parcels



Note. The same design is used with 9 item parcels from the STEU (three for anger, three for fear, and three for sadness)

1.2 Sources of Variation in Strategic EI: Verbal Ability or Interpreting Emotions?

Both the STEU and STEM are in a text-based format, where participants must read paragraphs of text before they undertake any emotional reasoning or knowledge. Thus, a vitally important concern is that the variation in test scores is due to EI rather than the ability to read the questions. Studies in both Chapter 3 and Chapter 5 examined aspects of this issue, but results are somewhat inconclusive. Results from Chapter 3 ascertained that the relationship between STEU, STEM, and emotion perception scores (as assessed by the text-based Stories test from the MEIS) was not due to vocabulary-related variance, indicating that the tests assess unique EI-related variation unrelated to verbal ability. However, that study also found that some EI scores correlated more highly with vocabulary than other EI scores, suggesting such EI tests measure verbal ability to a greater extent than EI. Chapter 5's results demonstrated that cultural differences in EI test scores were due to English language experience, implying that knowledge of the language is clearly related to scores on EI tests. In addition, research from Chapter 5 found that although EI was separable from Gc, the two constructs were very strongly related. As a whole, these results indicate that a considerable proportion of the variation on these text-based EI tests may be due to verbal ability, but equally that some of the variation is due to a unique EI construct. The basic challenge for research lies in partitioning the variation of these text-based EI tests into EI-related and reading-related variance, and then deciding if the EI-related proportion is large enough to posit EI as a separate entity. The current study examines this issue by considering the distinction between crystallized intelligence (Gc) and reading and writing ability (Grw).

McGrew's (1997) distinction between Gc and Grw can be used to test whether EI scores relate to intelligence because of the reading ability required to comprehend the question (Grw), or because emotional content constitutes another form of acculturated knowledge (Gc). In the current study, scores on a reading comprehension test act as a Grw control when examining relationships between Gf, Gc, Strategic EI, and emotion recognition ability (ERA; a set of nonverbal abilities conceptually equivalent to the Perception branch of the four-branch model). It is expected that

although both the STEU and STEM relate to reading comprehension, they should still relate meaningfully to each other when reading comprehension is accounted for. In addition, the relationship between EI and other measures of intelligence should not be entirely due to reading comprehension skill – if this were the case then the text-based format of the tests would be acting as a method effect. It is expected that both STEU and STEM should relate to Gf and Gc, and that these relationships will remain non-trivial (i.e., $r > .10$) when reading comprehension is accounted for, as this should not be the only source of shared variance between EI and other forms of intelligence.

1.3 Strategic EI and ERA: A Distinct EI Factor or Parts of Gc and Gf Respectively?

The contribution of reading skill to EI variance is further examined through the relationship between text-based measures of EI (the STEU and STEM) and non-verbal measures of EI (emotion recognition ability or ERA). ERA research tools come from the emotions research tradition rather than EI field and assess how well people can recognise emotions in different stimuli. In the current chapter, ERA is measured in both vocal expressions and facial expressions. These are assessed with the Recognition of Affect in a Foreign Language measure of vocal expression recognition (RAFL; Scherer et al., 2001), and a measure of facial expression recognition derived from the NIMSTIM stimulus set of facial expressions.[44] ERA tools are scored by standards based on emotions research rather than by consensus, and are conceptually linked to the Emotion Perception branch of the four branch hierarchical model of EI, although operationalised somewhat differently to the MEIS and MSCEIT Perception tests (Roberts et al., submitted).

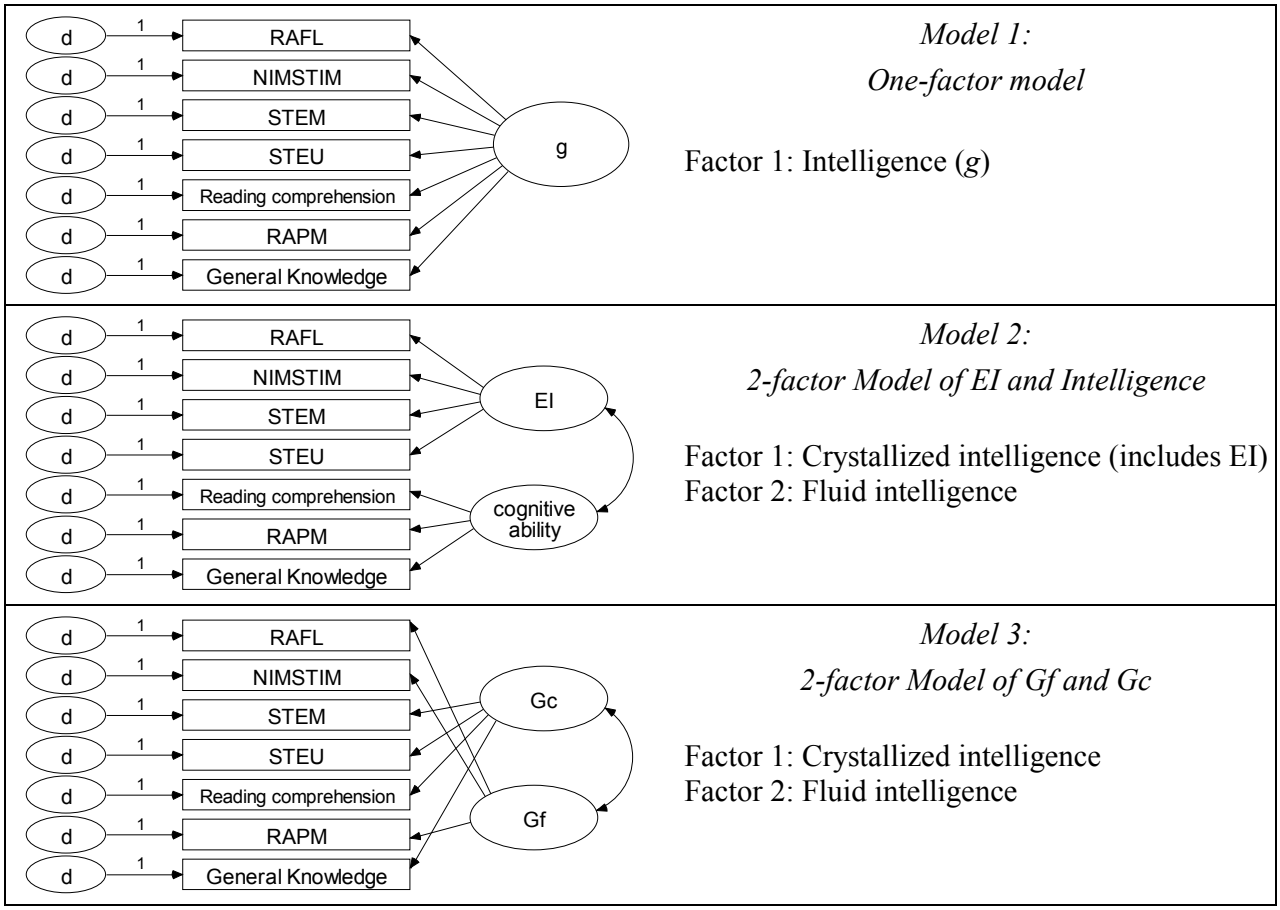
If EI is indeed a distinct type of intelligence, then different components of EI (i.e., Strategic EI and ERA) should relate strongly to each other. Also, if EI is a distinct type of intelligence, relationships should be stronger among EI measures than between EI and other intelligences. This idea will be tested by comparing a one-factor model (where Gf, Gc, Strategic EI, and ERA markers

[44] Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

all load on one factor – a *g* model) to a model distinguishing between EI and other forms of cognitive ability (a general EI versus *g* model). For the purposes of clarity, these one-factor and two-factor models are illustrated in Figure 6.2 as Model 1 and Model 2.

Figure 6.2

The Structure of Three Models of Intelligence to be Tested in this Chapter



Roberts et al. (submitted) found small to moderate relationships between Strategic EI (as measured by the MSCEIT) and two ERA measures: the JACBART (which assesses recognition of emotions in facial expressions) and the RAFL (which assesses recognition of emotions in vocal expressions, and is the same task used in the current study). However, in a factor analysis including MSCEIT subtests, Gf, Gc, and ERA markers, ERA tasks loaded on Gf and Gc factors rather than either Experiential or Strategic EI factors defined by MSCEIT tasks. Both ERA tasks loaded on Gf

and the RAFL also loaded on Gc. This evidence contradicts the idea that EI tasks will together form a distinct factor separable from other cognitive abilities. Roberts et al.'s analysis suggests that ERA is relatively distinct from Experiential EI (despite a strong conceptual relationship), and instead relates more strongly to Gf and Gc. In conjunction with results from Chapter 5, where Strategic EI was strongly related to Gc, this suggests that ERA and EI may form parts of Gf and Gc respectively, rather than both forming a single EI construct. For this reason, a third possible model was set up where ERA indicators load on a Gf factor and Strategic EI on a Gc factor (with the RAFL cross-loading, as was the case in Roberts et al.). This model is shown as Model 3 in Figure 6.2. In the current chapter, these three models will be compared with exploratory and confirmatory factor analysis.

1.4 Relationship of the Strategic EI and ERA to Personality

1.4.1 Relationship of the STEU and STEM to Personality

Relationships of the STEU and STEM to personality have been relatively consistent with findings for the MEIS/MSCEIT research tradition, with small relationships to Agreeableness and Openness (Brackett & Mayer, 2003; Brackett et al., 2004; Lopes et al., 2003, 2004; Roberts et al., 2001; Schulte et al., 2004). In Chapter 3, results indicated that the STEU and STEM related to Agreeableness but results from Chapter 5 isolated this relationship to the Morality facet of Agreeableness. The study outlined in Chapter 5 also found that all Strategic EI measures related to Openness, which was not the case for results obtained in Chapter 3. Results for Openness have also been a little inconsistent for the MSCEIT, with Lopes et al.'s (2003) negative correlation with Openness going against a general trend. It seems likely that the magnitude of the relationship of Strategic EI with Openness is small enough so that the margin of error may affect the results of any individual study. As such, replication is important.

This chapter briefly re-examines the relationship of the STEU and STEM to personality for two reasons. Firstly, the low reliability of the short forms of the STEU and STEM used in Chapter 5 meant that their relationship to the facets of Neuroticism may have been underestimated (which

effects the interpretation of Chapter 4's results relating STEU and STEM to *state* anxiety, depression, and stress). This chapter examines whether relationships to Neuroticism facets can be replicated with longer forms of the STEU and STEM, and also re-examines the STEU and STEM's relationship to the Morality facet of Agreeableness. Secondly, since both Chapter 4 and Chapter 5 used the STEM in multiple-choice format, the current chapter uses the STEM in ratings-based format, so as to get information on both possible formats. This may be important given that personality is assessed by a series of ratings, and thus ratings-based measures may relate more strongly to personality scales (although this increase in correlation was not significant in the research from Chapter 3). Thus, this chapter examines the relationship of STEU and STEM to the five factor model of personality, as well as key facets of Neuroticism (Anger, Anxiety, Depression, and Stress) and the Morality facet of Agreeableness (so chosen because of the high correlations with Morality in results from Chapter 5).

1.4.2 Relationship of the Two ERA Measures to Personality

In contrast to research on text-based measures of EI, few studies have examined the relationship between ERA and personality using standard, reliable personality measures (Matsumoto et al., 2000). In fact, there are no published studies relating vocal ERA to personality, and only two relating ERA in faces to the five-factor model dimensions of personality: those of Matsumoto et al. (2000) and Rubin, Munz and Bommer (2005). In a series of four studies, Matsumoto et al. found scores on the Japanese and Caucasian Brief Affect Recognition Test (JACBART) consistently showed a small degree of relationship to Openness (r = .21, .38 .31 and .33, $p < .05$ in all cases), and a smaller, less consistent relationship to Conscientiousness (r = .11, and .24, $p < .05$; and r = .15 and .24, $p > .05$). No significant relationships to Agreeableness were found across the five studies, and relationships to Extraversion and Neuroticism were found in only one of the four studies, which used a very small sample (N = 27; r = .47 and r = -.45 respectively). Rubin et al. found that Diagnostic Analysis of Nonverbal Accuracy (DANVA) scores were unrelated to Agreeableness and Extraversion, but did not examine other personality dimensions. In sum, relationships between ERA measures and personality seem similar to the relationships between intelligence and personality: The

only strong consistent relationship found is with Openness. However, no studies have examined the relationship between *vocal* ERA and personality, such that these results cannot be generalised across ERA but might hold for facial expression recognition only. The relationship between the five factor dimensions of personality and ERA in both faces and voices is examined in the current study. It is expected that ERA measures will relate to Openness, and possibly Conscientiousness, but not to any other personality dimensions.

1.5 Predicting University Grades: An Extension of Chapter 3

In Chapter 3, results indicated that the STEU (but not the multiple-choice STEM) had some incremental validity in predicting academic achievement over and above vocabulary and personality. The current study expands this analysis in two ways. Firstly, the study from Chapter 3 used vocabulary as the only measure of intelligence, whereas the current study measures intelligence more broadly, using three indicators representing different primary mental abilities (Gf, Gc, and Grw). Secondly, Chapter 3 only examined the possibility of incremental validity for the multiple-choice STEM, and this study will use the ratings-based STEM. Thirdly, several measures of academic achievement were used in the current study, collected both by self-report and from official university records.

Participants were undergraduates studying psychology in the second semester of their first year of their degree. Since most of them had taken psychology the previous semester, they were asked to report their mark in this subject, as well as their University Admissions Index (UAI). The UAI is a measure of secondary school achievement in New South Wales composed of coursework and examinations in the final year of secondary school. After the study was completed, participants' psychology grade at the end of the second semester was obtained from official records. Participants' average mark over the first year of university (their Weighted Average Mark or WAM) was also obtained from these official records.

Two separate regression models were conducted separately for these four indices of academic achievement: (1) UAI; (2) semester 1 psychology grade; (3) semester 2 psychology grade; and (4)

average grade at the end of first year (WAM). The first regression model predicted academic achievement from the EI score (either STEU or STEM) alone. The second regression model was a three-step hierarchical model, as in previous studies. In the first step, achievement was predicted by an intelligence composite. In the second step, the five factor dimensions of personality were added to the model. In the third step, EI scores were added to the model. Based on results from Chapter 3, and from Barchard's (2003) analysis of the MSCEIT, it seems likely that the STEU but not the STEM might relate to academic achievement, predicting a modest amount of variation in achievement beyond intelligence and personality.

1.6 Summary of Hypotheses

Hypothesis 1: Emotion specific models of the STEU and STEM differentiating between anger, fear, and sadness will provide a better fit to the data than one-factor models.

Hypothesis 2: STEU and STEM scores will relate moderately to strongly to each other (r = .40 to .60) even when reading comprehension is controlled, and will also relate non-trivially to Gf and Gc measures even when reading comprehension is controlled. These relationships are expected to be stronger for Gc than Gf.

Hypothesis 3: It is expected that Model 2 (a two-factor model with EI and cognitive ability as separate factors) will fit the data better than Model 1 (a one-factor model where all indicators load on the one cognitive ability factor). This would indicate that EI is distinct from other cognitive abilities. In exploratory research, Model 3 (where Strategic EI forms part of Gc and ERA forms part of Gf) will also be compared to the Models 1 and 2 to see if a Gf/Gc model can partition the variance in test scores in a meaningful way.

Hypothesis 4: Both Strategic EI and ERA measures will show a small degree of relationship to Openness (r = .30 or less), but only Strategic EI (i.e., STEU and STEM scores) will relate to Agreeableness, especially the Morality facet. ERA measures may also relate to Conscientiousness. In addition, STEU and STEM scores will relate only trivially to trait Anxiety, Depression, and Stress.

Hypothesis 5: EI will show some incremental validity in explaining academic achievement over and above the effects of intelligence and personality, and this effect will be stronger for STEU scores than STEM scores.

2. Method

2.1 Participants

118 first year undergraduate psychology students from the University of Sydney (68 female) participated in this experiment for course credit. Ages ranged from 18 to 32 with a median of 19 (M = 19.3, SD = 1.8). 20 participants were non-native English speakers. Participants self-reported the cultural groups they identified with. 54 (45.4%) identified with Australian or Anglo-Celtic cultures only, and culture was considered as a majority/minority dichotomy in all further analysis, with Australian/Anglo-Celtic as the majority. Cultural minority groups mostly identified as Asian cultures (25% of total) and European cultures (11% of total).

2.2 Materials

2.2.1 Emotional Intelligence (EI) Tests

Situational Test of Emotional Understanding (STEU). Participants were given 40 items from the STEU to complete at their own pace. 25 items were selected from the study in Chapter 3 for their high item-total correlation, and 15 additional items were created based on workplace and personal life analogs of the abstract items for fear, anger, and sadness (so that an examination of emotion specificity was possible). In this study, the STEU is scored according to standards derived from appraisal theory. Appendix 6.1 re-produces the additional 15 STEU items used in the current study.

Situational Test of Emotion Management (STEM). Participants were given 24 items from the STEM to answer in ratings-based format (8 items related to sadness, 8 to anger, and 8 to fear).[45] Items were selected from the study in Chapter 3 for their high item-total correlations. In this study, the

[45] Due to a computer error, responses to one of the sadness-management items were not recorded, effectively making this a 23-item test.

STEM was scored by standardised expert distance (i.e., the same scoring rubric used for the ratings-based STEM in Chapter 3, as this was the most reliable). STEM items used in this study are marked with an asterisk in Appendix 2.8.

2.2.2 Emotion Recognition Ability (ERA) Tests

Recognition of Affect in a Foreign Language (RAFL). The RAFL (Scherer et al., 2001) consists of 30 items assessing recognition of emotion in tone-of-voice. In each item, participants listen to a nonsense phrase given twice and then need to select which of five emotions the voice expresses (joy, sadness, anger, fear or no emotion). The task was computerised and participants were each given a set of headphones. The RAFL has been used frequently in cross-cultural research on emotion (Scherer et al., 2001). Internal consistency reliability has been reported at .51 (Roberts et al., submitted).

NIMSTIM Faces Test. The NIMSTIM Faces Test was constructed specifically for use in the current study, using stimuli from the NIMSTIM stimulus set (Tottenham, 2004). 48 emotion-expressing faces of African American, Caucasian, and Asian appearance (12 for each emotion of sadness, anger, fear, and joy) were presented with a forward and backwards mask of the same face with a neutral expression. The forward mask was presented for one second before the stimulus face and the backwards mask was left onscreen until participants selected a response from four possible options (sadness, anger, fear, and joy). After a response was selected, the task moved immediately to presentation of the next forward mask. Half of the stimuli faces were presented for 200 msec and half for 100 msec, in line with findings from Matsumoto et al. (2000) on the reliability resulting from different presentation latencies. However, these presentation times were probably too long, as there was 100% accuracy on 6 items (5 of which were joy items) and greater than 95% on 15 more items (7 of which were joy items). These 21 items were excluded before computing total scores, to create more variability in the test. The final score was thus derived from 27 items, none of which pertained to joy-recognition.

2.2.3 Intelligence Tests

Raven's Advanced Progressive Matrices. Participants were given a subset of 20 items from Raven's Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 1998) to complete within an 8 minute time limit.[46] Raven's Progressive Matrices is considered a pure measure of fluid intelligence (Gf; Carroll, 1993), although the task also involves some spatial abilities (Gv; e.g., Colom et al., 2004). In each item, a three by three grid of designs follows a pattern both across and down the grid and participants must select from eight possible responses the pattern that goes in the lower right corner to complete both patterns.

Knowledge. Participants were given set 1 (28 items) from the Knowledge test from the Intelligence Structure Test (I-S-T, Amthauer, Brocke, Leipmann, & Beauducel, 2000; translated into English with culturally bound items modified for an Australian population by Danthiir, Wilhelm, & Schacht, submitted). Items assess general knowledge in several domains, and participants must select the best answer from five possible options. General knowledge is known to be a marker of crystallized intelligence (Gc; Carroll, 1993).

Reading Comprehension Test. Participants were given a subset of 43 items from Part II of the Co-operative Reading Comprehension Test, Form Y (ACER, 1978) with an 8 minute time limit. Participants were required to read excerpts of text and then answer multiple-choice questions that assessed their understanding of these excerpts. Reading comprehension is a measure of crystallized intelligence or acculturated knowledge (Gc) under Carroll (1993) and a measure of reading and writing (Grw) under McGrew (1997). In the current study, reading comprehension will be used as a Grw control when examining relationships between Strategic EI, ERA, Gf, and Gc, but is expected to load on intelligence and Gc factors in structural analyses of these test scores.

---

[46] Items for this 20 item short form of the RAPM were selected by Birney & Bowman (in preparation).

2.2.4 Personality Tests

OCEANIC-20. Participants completed the 20-item version of the OCEANIC, a measure of the five factor model of personality (Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed; Roberts, 2000). Reliabilities from Chapter 3 ranged from .71 (for Openness) to .85 (for Extraversion) when participants were restricted to the first five points of the scale. In the current study, participants are instructed to rate their agreement with 20 statements (4 for each of the five factors of personality) on a scale of 1 (Never) to 6 (Always).

Scales from the IPIP NEO-PI-R. Five facet scales from the short IPIP version of the NEO-PI-R (Johnson, 2004) that correlated with the STEM and STEU in Chapter 4 ere included in the current study to assess their relationship with the ratings-based version of the STEM, and replicate Chapter 4's relationships between these scales and the STEU. These scales were Anger, Anxiety, Depression, and Stress (all facets of Neuroticism) and Morality (a facet of Agreeableness).

2.3 Procedure

Participants volunteered to participate based on a brief description of the experiment. After reading a 2-page information form, participants signed a consent form indicating their willingness to participate in the experiment and whether or not they gave the experimenter permission to access their grades for that semester (113 participants agreed their grades could be accessed, and 94 self-reported their psychology grade from the previous semester). After signing consent forms, participants completed all tests on PC computers. Test order was counter-balanced across four conditions, to ensure that fatigue or boredom related factors were counter-balanced across tests. Half the participants completed ERA tests followed by other tests, and the other half completed the ERA tests last. The other tests (EI, intelligence and personality) were given in two possible orders: (1) STEM, demographic questions, RAPM, STEU, Knowledge, Reading Comprehension, OCEANIC; and (2) STEU, demographic questions, reading comprehension, STEM, Knowledge, RAPM, and OCEANIC. This study was approved by the Sydney University Human Research Ethics Committee.

2.4 Analyses

Analysis 1: Reliability and Descriptive Statistics. Reliability and descriptive statistics were calculated and where possible compared to previous research to make sure that test scores were functioning as expected. Gender differences were also calculated.

Analysis 2: Structural Models of the STEU and STEM Examining Emotion Specificity. Structural models of STEU and STEM item parcels outlined in Figure 6.1 were compared to examine whether understanding and management of emotions might be distinct for different types of emotions.

Analysis 3: Relationships between Strategic EI, Gf, and Gc after controlling for Reading Comprehension. The relationships between Strategic EI scores, Gf, and Gc indicators were calculated with and without controlling for levels of reading comprehension to assess how much shared variability was due purely to the reading ability needed to answer the text-based EI items.

Analysis 4: Relationships between EI and Personality. Zero-order correlations of the STEU, STEM, RAFL, and NIMSTIM Faces with the five factor model dimensions of personality and with trait Anxiety, Depression, and Vulnerability to Stress (as well as Morality and Anger) were calculated.

Analysis 5: Structural Models of the Distinction between EI (Strategic EI and ERA) and Other Cognitive Abilities. Exploratory factor analysis of Strategic EI, ERA, and Intelligence (Gc, Gf, and Grw) indicators tested whether EI and cognitive ability form part of the same construct. Confirmatory analyses then test whether a one-factor model, a two-factor EI versus cognitive ability model, or a two-factor Gf versus Gc model best explained the relationship between EI, ERA and more traditional cognitive abilities.

Analysis 6: Incremental Prediction of Academic Achievement by Strategic EI. Two separate sets of regressions tested whether STEU and STEM scores predicted academic achievement. Simple linear regressions predicted whether STEU and STEM scores predict academic achievement at all, and a three-step hierarchical model tests whether STEU and STEM scores predicted achievement after controlling for intelligence and personality. Four different measures of academic achievement

were used: (1) self-reported psychology grade from previous semester; (2) psychology grade for current semester (drawn from official records); (3) average university grade at the end of the first year of study; and (4) University Admissions Index (UAI). Regressions were run separately for STEU and STEM scores for each measure of academic achievement.

3. Results

3.1 Reliability and Descriptive Statistics

3.1.1 Reliability and Descriptive Statistics for Cognitive Ability, Strategic EI and ERA

Reliability and descriptive statistics are shown in Table 6.1, along with comparisons to other data sets (where possible) and comparisons of male and female performance. The low mean for the reading comprehension was due to lack of people finishing all items (only 10 participants completed the last question of the test; i.e., speed as well as accuracy contributed to the score). The I-S-T Knowledge test, STEU, and STEM were appropriately reliable for research purposes, but there were some problems with the reliability of the RAFL and the NIMSTIM Faces. Even the 21 NIMSTIM items that showed greater than 95% accuracy (6 of which showed 100% accuracy) were excluded, reliability was still only .56. The RAFL had a reliability of .43, and 8 items were removed to improve reliability (using half the sample was to select items and the other half to calculate reliability). However, descriptive statistics were also reported for the full RAFL in order to compare results with previous analyses. Mean scores on the STEU, STEM, and RAFL were consistent with previous analyses. Although reliability estimates for the STEM scores were lower in the current study than reported in Chapter 3's study, this is likely to be because the data from Chapter 3's study was used to select the items (and therefore the higher reliability is a statistical artefact). The slightly higher mean of the current sample might also be due to this issue, but might also be due to differences in the sample (in Chapter 3, half of the students completing the ratings-based STEM were from a rural rather than city campus, where academic requirements for entry were less stringent).

There was a significant male advantage on the Knowledge test, consistent with previous studies (e.g., Ackerman, et al., 2001; Colom et al., 2004; Lynn et al., 2001). As in previous chapters there was a significant female advantage on the STEM, echoing the results from Chapter 5.

Table 6.1

Reliability and Descriptive Statistics: Comparison to Other Research and Gender Differences

	Whole sample (N=118)			*Comparison Samples*				*Males (N=50)*		*Females (N=68)*		*Gender d*
	α	*M*	*SD*	*α*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*Sex*
Knowledge	.64	0.58	0.13	-	-	-	-	0.65	0.11	0.54	0.12	0.87**
RAPM	-	0.54	0.17	-	-	-	-	0.56	0.15	0.53	0.18	0.16
RC	-	0.26	0.11	-	-	-	-	0.27	0.11	0.25	0.12	0.21
STEU (all)	.70	0.71	0.11	-	-	-	-	0.70	0.11	0.72	0.11	-0.21
STEU (subset) [a]	.61	0.72	0.17	.72	.72	.21	0.00	-	-	-	-	-
STEM	.74	0.72	0.20	.86	.63	.29	0.37	0.67	0.19	0.76	0.20	-0.44*
RAFL [b]	.43	0.67	0.10	.51	.66	.10	0.10	0.66	0.09	0.69	0.09	-0.32
RAFL (short)	.56	0.70	0.12	-	-	-	-	0.68	0.12	0.71	0.12	-0.25
Faces (short)	.67	0.88	0.11	-	-	-	-	0.87	0.11	0.88	0.10	-0.05
O [c]	.80	16.35	4.32	.74	14.88	3.79	0.36	17.98	3.72	15.18	4.39	0.68**
C	.79	15.19	3.99	.73	15.10	3.56	0.02	14.38	3.71	15.82	4.13	-0.36
E	.81	16.29	4.25	.80	15.02	3.69	0.32	14.12	4.54	14.44	4.07	-0.08
A	.58	19.44	2.65	.79	19.45	3.05	0.00	18.00	2.65	18.76	2.62	-0.29
N	.82	14.34	4.74	.65	13.72	3.33	0.15	12.16	4.61	15.94	4.22	-0.86**
Anxiety [d]	.84	12.72	4.64	.87	12.79	4.41	-0.02	10.54	4.28	14.26	4.28	-0.87**
Anger	.81	12.34	4.23	.78	14.08	3.99	-0.42	11.72	4.75	12.91	3.66	-0.29
Depression	.81	11.34	4.00	.84	10.68	4.14	0.16	10.86	4.04	11.72	3.98	-0.21
Vulnerability	.80	11.74	4.11	.74	12.85	3.83	-0.28	10.12	4.01	12.94	3.82	-0.72**
Morality	.76	19.24	3.28	.73	19.55	3.08	-0.10	18.44	3.52	19.81	3.01	-0.42*
UAI	-	90.07	7.50	-	88.23	11.98	0.18					
Psych S1	-	63.54	11.27	-	68.03	11.67	-0.39					
Psych S2	-	61.98	10.43	-								
WAM	-	61.66	10.14	-	66.76	13.40	-0.43					

Note. Knowledge = General Knowledge; RAPM = Raven's Advanced Progressive Matrices; RC = Reading Comprehension; STEU (all) = all 40 STEU items; STEU (subset) = 25 items common to this study and Chapter 3; RAFL = Recognition of Affect in a Foreign Language (all 30 items); RAFL (short) = 22 most reliable items from the RAFL; Faces (short) = 27 NIMSTIM Faces items with difficulties < 0.95; O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism.

[a]Comparison sample for the STEU and STEM from Chapter 3. [b]Comparison sample for the RAFL from Roberts et al. (submitted). [c]Comparison sample for the OCEANIC from MacCann et al. (2004). [c]Comparison sample for Anxiety, Anger, Depression, Vulnerability and Stress from Chapter 5.

3.1.2 Reliability and Descriptive Statistics for Personality

All personality dimensions were adequately reliable for present purposes, although Agreeableness was a little low, with an index of .58. There were some small differences in personality compared to previous studies: The current sample was more Open to Experience, more Extraverted, and had lower trait Anger. However, all effect sizes were smaller than .50, the cut-off for a "medium" effect (Cohen, 1988). Gender differences in personality were consistent with Chapters 3 and 5, as well as other published research. Males scored higher on Openness, and females scored higher on Neuroticism, including Neuroticism facets of Anxiety and Vulnerability to Stress. Females did *not* score significantly higher than males on Agreeableness, but did score significantly higher on the Morality facet of Agreeableness.

3.6 Emotion Specificity of the STEU and STEM

To check whether Understanding and Management constitute different constructs for different emotions, structural analysis of STEU and STEM item parcels were conducted separately for the STEU and STEM. Each analysis compared a one-factor model (where sadness-, anger-, and fear-related items all loaded on the same factor) to a three factor model (where sadness-, anger-, and fear-related items formed three separate factors; see Figure 6.1 for diagrams of these structural equation models). Although the current sample is rather small for structural equation modelling (list-wise N=109), results suggest directions for future research (i.e., replication is certainly necessary).

Descriptive statistics and correlations between the items parcels are provided in Table 6.2 so that the interested reader can explore the structural analysis further. Parameters for the STEU are shown on the left and lower left triangle and parameters for the STEM are shown on the right and upper right triangle. It can be seen that the correlations are stronger for the STEM than the STEU.

Table 6.2

Descriptive Statistics and Correlations between Nine STEU Item Parcels (Left and Lower Left) and Nine STEM Item Parcels (Italicised in Right and Upper Right)

	STEU	*Correlations*									*STEM*
	Mean (SD)	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Mean (SD)*
1. Anger subset 1	1.33 (.86)		*.16*	*.21*	*.18*	*.19*	*.19*	*.10*	*.17*	*.06*	*3.89 (.78)*
2. Anger subset 2	2.31 (.80)	.17		*.34*	*.08*	*.22*	*.26*	*.17*	*.22*	*.31*	*4.01 (.80)*
3. Anger subset 3	1.65 (.48)	-.02	.11		*.14*	*.26*	*.27*	*.17*	*.23*	*.36*	*2.47 (.85)*
4. Fear subset 1	1.93 (.68)	.06	.14	.05		*.36*	*.41*	*.24*	*.19*	*.04*	*4.09 (.99)*
5. Fear subset 2	2.07 (.69)	.12	.04	.02	.03		*.48*	*.33*	*.33*	*.36*	*3.88 (1.11)*
6. Fear subset 3	1.49 (.48)	.06	.20	.05	.27	.00		*.19*	*.29*	*.21*	*2.60 (.94)*
7. Sadness subset 1	2.62 (.55)	.20	.27	.12	.17	.15	.04		*.28*	*.24*	*3.96 (.80)*
8. Sadness subset 2	1.49 (.52)	-.03	.10	.06	.06	.16	.14	.14		*.29*	*2.53 (.96)*
9. Sadness subset 3	1.89 (.36)	.12	.18	.07	.14	.13	.10	.29	.29		*2.69 (.86)*

Note. At $r = .19$, $p < .05$; at $r = .25$, $p < .01$; missing data deleted list-wise, N=109.

Factor loadings, factor correlations and fit indices for the structural analysis of STEU item parcels are shown in Table 6.3. For the analysis of the STEU, Model 2 (the three-factor model) fit significantly better than Model 1 (the one-factor model), showing a significant reduction in χ^2 values and a large decrease in the AIC indices. Fit indices were acceptable for Model 2 (i.e., RMSEA < .06 and both CFI and GFI > .95). However, several of the paths in this model were not significant, and the model overfit the data ($\chi^2 < df$). Thus this evidence for emotion-specificity in the Understanding construct needs to be interpreted with caution. Fear and Anger factors were more strongly correlated with each other ($r = .74$) than with Sadness ($r = .58$ and $r = .47$ respectively).

Table 6.3 also shows the factor loadings, factor correlations and fit indices for one-factor and three-factor models of the STEM item parcels. An emotion-specific model for Management fit the data better than a general Management factor, according to both χ^2 and AIC differences. However, fit indices were good for the one-factor model as well as the three-factor model (RMSEA < .06 and both CFI > .95). Although the three-factor model was overfitted, all paths in the model were significant for

the STEM analysis, such that the three-factor model appears to be the best fit to the data. Anger, Fear and Sadness factors were highly inter-correlated (all correlations were greater than .60) indicating considerable similarity between these constructs.

Table 6.3

Factor Loadings, Factor Correlations, and Fit Indices for the Structural Analysis of Nine STEU Item Parcels for Five Different Models

	Structural Analysis of the STEU				*Structural Analysis of the STEM*			
	Model 1	*Model 2*			*Model 1*	*Model 2*		
	(27 df)	*(24 df)*			*(27 df)*	*(24 df)*		
		Factor 1 Anger	*Factor 2 Fear*	*Factor 3 Sadness*		*Factor 1 Anger*	*Factor 2 Fear*	*Factor 3 Sadness*
1. Anger 1	.30	.28#			.30	.33		
2. Anger 2	.38	.54			.42	.54		
3. Anger 3	.43	.19#			.47	.62		
4. Fear 1	.60		.56		.44		.55	
5. Fear 2	.54		.35		.50		.55	
6. Fear 3	.55		.56		.48		.46	
7. Sadness 1	.46			.52	.45			.50
8. Sadness 2	.21#			.12#	.70			.74
9. Sadness 3	.59			.48#	.63			.67
Correlations								
Fear (F2)		.74				.79		
Sadness (F3)		.58	.47			.61	.76	
Fit Indices								
χ^2	47.51		16.89		33.07		20.39	
$\Delta\chi^2$			30.63*				12.68**	
GFI	.92		.97		0.94		0.96	
CFI	.84		1.00		0.96		1.00	
RMSEA	.08		0.00		0.04		0.00	
AIC	83.51		58.89		69.07		62.39	

Note. All RMSEAs are not significantly greater than .05. All paths in the STEM analysis are significant at $p < .01$

#These paths not significant at $p < .05$

*$p < .05$, **$p < .01$

Table 6.4 shows the correlations between factor scores for STEU Anger, Sadness and Fear, and STEM Anger, Sadness and Fear. All are small (.10 to .33) and there seems to be no indication that correlations for scores involving the same emotion (e.g., Fear-Management with Fear-Understanding) are any stronger than correlations for scores involving different emotions (e.g., Anger-Management with Fear-Understanding). Although STEM-Fear correlates more strongly with STEU-Fear than with STEU-Anger or STEU-Sadness (and vice-versa), STEM-Sadness also correlates more strongly with STEU-Fear (than STEU Anger or Sadness).

Table 6.4

Correlations between Anger- Fear- and Sadness-Specific Factor Scores for the STEU and STEM

	STEM Anger	*STEM Fear*	*STEM Sadness*
STEU Anger	-.17	-.25**	-.10
STEU Fear	-.17	-.33**	-.20*
STEU Sadness	-.31**	-.20*	-.14

3.2 Relationships between Strategic EI, ERA, Gf and Gc when Controlling Grw

Correlations between Intelligence, Strategic EI, ERA, and personality are shown in Table 6.5. Partial correlations between STEU, STEM, RAPM, Knowledge, and Faces when Reading Comprehension is controlled are given in italics, showing whether EI tests still relate to each other, Gf, and Gc after accounting for reading ability. In addition, relationships of the emotion specific factor scores for the STEU and STEM to intelligence and personality are also presented in this table to ascertain whether scores for different emotions have different correlates.

Table 6.5

Correlations between Intelligence, EI, ERA, and Personality (Partial Correlations Controlling for Reading Comprehension Shown in Italics)

	1	2	3	4	5	6	7	4A	4F	4S	5A	5F	5S
1. Reading								.13	.19	.09	.13	.22	.08
2. RAPM	.33		*.25*	*.27*	*.16*	*.04*	*.25*	.14	.16	.12	.10	.16	.23
3. Knowledge	.43	.35		*.17*	*.09*	*-.08*	*.02*	.19	.11	-.01	-.01	.16	.08
4. STEU	.35	.35	.29		*.41*	*.23*	*.16*	.55	.68	.47	.30	.44	.28
5. STEM	.17	.20	.16	.43		*.24*	*.11*	.22	.30	.26	.72	.82	.80
6. RAFL	.30	.13	.06	.31	.27		*-.01*	.13	.24	.15	.13	.33	.16
7. Faces	.13	.28	.07	.20	.13	.03		.03	.13	.42	.13	.14	.05
8. O	.33	.17	.34	.12	.07	.19	-.02	.00	.06	-.16	-.10	.08	.00
9. C	-.18	-.05	-.31	-.09	.11	-.12	.06	-.10	-.07	-.09	.08	.06	.07
10. E	.07	.01	.01	.10	.02	-.06	.03	.12	.11	.01	.00	.04	.00
11. A	.01	.03	-.08	.19	.22	.22	.25	.09	.21	.26	.18	.23	.18
12. N	.04	-.11	-.22	-.05	.11	.12	.10	-.13	.05	.08	.16	.11	.01
13. Anxiety	-.02	-.06	-.11	-.11	.00	.09	.04	-.12	-.01	.06	.13	.05	.03
14. Anger	.02	-.05	-.37	-.06	.08	.06	.10	-.14	-.04	-.04	-.07	-.02	.09
15 Depression	.10	.00	.04	-.03	.01	.08	.02	-.03	.05	-.04	.01	.00	.00
16. Stress	-.05	-.04	-.33	-.08	.08	.04	.08	-.04	.02	.08	.13	.03	.06
17. Morality	-.10	.08	.03	.25	.28	.15	.11	.21	.27	.25	.22	.25	.17

Note. 4A = STEU Anger factor; 4F = STEU Fear factor; 4S = STEU Sadness factor; 5A = STEM Anger factor; 5F = STEM Fear factor; 5S = STEM Sadness factor. Missing data was deleted list-wise, $N = 109$; When $r > .24$, $p < .01$; when $r > .19$, $p < .05$

All correlations between Intelligence, EI, and ERA are positive, indicating that EI and ERA are measuring intelligence to some extent. STEU and STEM scores were moderately related, consistent with results from Chapter 3 ($r = .43$ in the current study compared to $r = .49$ in Chapter 3). This relationship declined only very slightly after reading comprehension was accounted for,

indicating that the correlation between STEU and STEM is clearly not due to just to the skills involved in comprehending the text of the items. Relationships to ERA measures were similarly small but non-trivial for both the STEU and STEM both before and after reading comprehension was controlled.

The STEU correlated positively and non-trivially to RAPM and Knowledge both before and after Reading Comprehension was controlled for (although the partial correlation with Knowledge was not significant). This indicates that relationships between STEU scores and intelligence are not due only to the reading skill required to answer STEU items. The STEM was significantly related to RAPM but not Knowledge before controlling for Reading Comprehension (although the relationship to Knowledge was non-trivial). Both these relationships declined to trivial non-significance when reading comprehension was controlled. That is, relationships between STEM scores and intelligence may be partly due to the reading skill required to comprehend STEM items.

The ERA measures were unrelated to each other. Even examining ERA separately for each emotion (i.e., sadness-recognition, fear-recognition, anger-recognition etc in both faces and voices) produced no significant relationships between Faces and RAFL. This is similar to Roberts et al.'s (submitted) findings that the RAFL was only weakly related to a different facial expression recognition task (r = .22). Correlates of the emotion-specific factors of the STEU and STEM were not markedly different for the different factors, except for the NIMSTIM Faces' relationship to the STEU. Only the sadness factor of the STEU related to the NIMSTIM Faces. However, given that 90 possible correlations were examined, this could easily be a Type I error rather than a genuine difference, particularly given the low reliability of the NIMSTIM Faces.

3.3 Relationships of Strategic EI and ERA to Personality

Openness related positively to both Gc-type measures (Reading Comprehension and Knowledge), as would be expected from published findings (Ackerman & Heggestad, 1997). However, neither STEU nor STEM scores related to Openness in this study, consistent with results from Chapter 3, but in contrast to results from Chapter 5. These contradictory findings may be due to

the different personality instruments used in different studies (the OCEANIC in Chapter 3 and the current study, and the IPIP NEO-PI-R in Chapter 5). The RAFL showed a small but significant relationship to Openness, consistent with the hypothesis that ERA measures would relate to Openness. However, NIMSTIM Faces scores did not relate to Openness, in contrast to Matsumoto et al.'s (2002) findings for facial expression recognition when assessed with the JACBART. Also in contrast to Matsumoto et al.'s findings, neither measure of ERA related to Conscientiousness.

Strategic EI scores were significantly related to Agreeableness, consistent with Chapter 3. The STEU and STEM also related to the Morality facet of Agreeableness, consistent with Chapter 5. Both measures of ERA (the RAFL and NIMSTIM Faces) also showed significant positive relationships with Agreeableness (but not Morality), in contrast to Matsumoto et al. (2000) and Rubin et al. (2005). All EI measures (both Strategic EI and ERA) are thus related to positive social functioning to some extent, although they are unrelated to negative emotionality (i.e., Neuroticism) at either the broad or facet level. This replicates results from Chapter 5, indicating that the STEU and STEM's prediction of *state* anxiety, depression, and stress (found in the study in Chapter 4) was not due to overlap with *trait* Anxiety, Depression, and Stress. As in Chapters 3 and 5, no correlations with personality were high enough to indicate that EI (as defined by performance-based measures) overlaps too highly with constructs comprising the personality domain.

Females scored higher on both Morality and the STEM, and Morality and STEM scores were significantly correlated, meaning that gender differences on the STEM might be due to the STEM's relationship to Morality. In a replication of the study outlined in Chapter 5, an ANCOVA predicting STEM scores while controlling for Morality still found female superiority on the STEM ($F_{1,114} = 6.712$, $p = .011$). Marginal means on the STEM were $M = .68$ for men ($SE = .03$) and $M = .76$ for women ($SE = .02$).

3.4 Locating EI in the Structure of Intelligence: Exploratory and Confirmatory Factor Analysis

Exploratory factor analysis was run on the three cognitive ability, two Strategic EI and two ERA tests to inform the confirmatory factor analyses planned and presented in Table 6.6. Although

three eigenvalues were greater than 1, a three-factor solution resulted in a singlet for reading comprehension so a two-factor solution was instead extracted (using Maximum Likelihood and oblique rotation since factors were assumed to be correlated). Factor loadings are shown in Table 6.6. The pattern of loadings suggests that the two factors resemble an EI factor composed of STEU, STEM, and RAFL scores, and an Intelligence factor composed of Reading Comprehension, RAPM, and Knowledge scores. Faces did not load saliently on either factor, although had a stronger loading on EI than Intelligence.

This solution was checked with confirmatory factor analysis, and two other competing structures (described in Figure 6.2) were also examined. In Model 1, all 7 indicators load on one factor; whereas Model 2 distinguishes between an EI factor (STEU, STEM, RAFL, and Faces) and an Intelligence factor (RAPM, Reading Comprehension and Knowledge). Model 3 delineates the indicators into Gf (RAPM, NIMSTIM Faces, and RAFL) and Gc (Reading Comprehension, Knowledge, RAFL, STEU, and STEM). Discrepancy matrices and modification indices suggested that the RAFL loaded only on Gc and not on Gf, so the Gf cross-loading was excluded. Factor loadings and inter-correlations as well as selected indices of model fit are shown in Table 6.6 for all three models.

Fit indices are only good for Model 2, with CFI and GFI > .95 and RMSEA < .05. Fit indices for Model 1 were unacceptable, and indices for Model 3 were marginally acceptable (RMSEA not significantly greater than .05, and both CFI and GFI > .90 but not > .95). Model 2 (EI versus intelligence) fit the data significantly better than Model 1 (the one-factor model; $\Delta\chi^2(1) = 10.486$, $p < .001$; $\Delta AIC = 8.486$), indicating a distinction between EI and other cognitive abilities. Model 3 (Gf and Gc factors) did *not* show a significant improvement in fit over a one-factor model ($\Delta\chi^2(1) = 3.421$, $p > .05$; $\Delta AIC = 1.421$). In conjunction with the exploratory factor analysis supporting Model 2, these analyses suggest that Strategic EI and ERA abilities may cohere to form an EI construct that is distinguishable from other forms of intelligence. However, the correlation

between EI and Cognitive Ability factors was quite high (r = .66), echoing the high correlation between Strategic EI and Gc in Chapter 5.

Table 6.6

Exploratory and Confirmatory Factor Analysis of Cognitive Ability, Strategic EI and ERA Markers

	Exploratory Factor Analysis		*Confirmatory Factor Analyses*				
		(8 df)	*Model 1 (14 df)*	*Model 2 (13 df)*		*Model 3 (13 df)*	
			g	*EI*	*g*	*Gc*	*Gf*
Factor Loadings							
RC	.127	**.547**	.582	.655		.586	
RAPM	.169	**.446**	.498	.567			.797
Know	-.126	**.770**	.547	.608		.500	
STEU	**.655**	.156	.683		.809	.688	
STEM	**.613**	-.058	.469		.530	.474	
RAFL	**.460**	-.026	.388		.402	.399	
Face	.186	.114	.281		.252		.354
Factor r		.517			.658		.657
Fit Indices							
CFI			.886		0.983		0.911
GFI			.936		0.965		0.943
RMSEA			.086		0.035		0.079
AIC			53.185		44.699		51.764
χ^2	10.520		25.185		14.699		21.764

Note. Missing data was deleted list-wise, N = 109. All loadings are significant at $p < .05$. All RMSEAs are not significantly different from 0.05 at $p > .05$; and all χ^2s non-significant at $p > .05$.

Separate analyses of ERA indicators (RAFL and Faces) with intelligence and Strategic EI indicators (STEU and STEM) with intelligence proved informative here (these analyses are reproduced in Appendix 6.2). In a one-factor model of Strategic EI and cognitive ability, modification indices and discrepancy matrices suggested that STEU and STEM scores should form a separate

factor (i.e., factors were Strategic EI and cognitive ability, correlated at $r = .59$). In a one-factor model of ERA and intelligence, modification indices and discrepancy matrices suggested that RAPM and Faces should form a separate factor (i.e., factors were Gf and Gc, correlated at $r = .53$). That is, ERA scores fit into the Gf/Gc theory of intelligence (with auditory judgments of tone-of-voice loading with Gc and the visual inspection time task of facial expression recognition loading with Gf) but Strategic EI scores formed a different factor to intelligence.

3.5 Hierarchical Regressions of Academic Achievement onto EI

The incremental validity of EI in predicting academic achievement was examined separately for the STEU and STEM, and for four academic achievement indices: (1) self-reported University Admissions Index (UAI; a measure of secondary school performance based on state-wide exams as well as school-based tests and assignments); (2) self-reported semester 1 psychology grade; (3) semester 2 psychology grade (taken from official records); and (4) weighted average mark (WAM) over all university subjects for the first year of study (taken from official records). As in Chapters 3 and 5, two regressions were undertaken in each case: (1) a simple regression where achievement was predicted from EI alone; and (2) a three-step hierarchical regression. In the hierarchical regression, the predictors entered in each step were: (1) intelligence, composed of the first principal component of Reading Comprehension, Knowledge and RAPM; (2) personality (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism); and (3) EI (either STEU or STEM scores). The squared multiple correlations for the simple regression and each step of the multiple regression are shown in Table 6.7. For reasons of space, the standardized regression coefficients are given in Appendix 6.3 rather than in the table.

The STEM did not significantly predict any of the measures of academic achievement, even before intelligence and personality were accounted for, consistent with results from Chapter 3 for the WAM but not for psychology grades. STEU scores also did not significantly predict the WAM, even before accounting for intelligence and personality, in contrast to results from Chapter 3. Although STEU scores did predict UAI, and both sets of psychology grades, the proportion of variance

explained was low (5-6%), even before accounting for intelligence and personality. This is in strong contrast to the results from Chapter 3, where STEU scores accounted for 18% of the variation in psychology grade. After accounting for intelligence and personality, neither the STEU nor the STEM predicted any of the four indices of academic achievement.

Table 6.7

Regression Models Predicting Four Indices of Academic Achievement from a Combination of Intelligence (in Step 1), Personality (in Step 2) and EI (in Step 3), and from EI Only

	Step 1	*Step 2*	*Step 3*	
	Intelligence	*Personality*	*STEU*	*STEM*
Predicting University Admissions Index (UAI; N = 110)				
R^2	.32**	.36**	.36**	.36**
ΔR^2	.32**	.03	.00	.00
R^2 *(EI only)*[a]			.06*	.03
Predicting Self-Reported Psychology Grade (Semester 1; N = 93)				
R^2	.30**	.39**	.39**	.40**
ΔR^2	.30**	.08*	.01	.01
R^2 *(EI only)*[a]			.06*	.04
Predicting Future Psychology Grade (from Official Records, Semester; N = 110)				
R^2	.38**	.40**	.40**	.41**
ΔR^2	.38**	.02	.00	.01
R^2 *(EI only)*[a]			.05*	.03
Predicting Weighted Average Mark (WAM) for First Year (from Official Records; N = 110)				
R^2	.32**	.38**	.39**	.40**
ΔR^2	.32**	.07	.01	.01
R^2 *(EI only)*[a]			.04	.01

Note. Intelligence refers to the first principal component in an analysis of Knowledge, RAPM, and Reading Comprehension; Personality refers to the dimensions from the five-factor model (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism).

[a] R^2 *(EI only)* refers to the squared zero-order correlation between the criterion variables and EI (i.e., vocabulary and personality are not accounted for)

*$p < .05$, **$p < .01$.

4. Discussion

Results from this study highlight the ways that scores on emotion-related tests might segment differently according to the specific emotions dealt with as well as the processes involved in such tests. Relationships to Gf and Gc do seem to differ depending on the processes involved in the tasks, with inspection time paradigms relating to Gf, and auditory and knowledge-based processes relating to Gc. A re-examination of EI's relationships with personality and overlap with verbal ability supports the findings from previous chapters, as well as studies involving the MEIS/MSCEIT measurement paradigm. However, neither the STEU nor STEM incrementally predicted academic achievement in the current study. These issues are further discussed below.

4.1 Emotion Specificity of the STEU and STEM

The possibility that different constructs within EI might differ for different emotions was supported by analyses of the STEU and STEM. For both tests, emotion-specific models positing different factors for anger, fear, and sadness provided a better fit to the data than one Understanding or Management factor (although emotion-specific factors were strongly correlated with each other in both cases). However, processing of the specific emotions (fear, anger, and sadness) did not appear to be the same across Understanding and Management (e.g., understanding sadness was no more related to managing sadness than to managing fear or anger). Both these findings may be important when designing future research to ascertain the potential applications for EI tests.

If understanding (or managing) emotions is specific to different types of emotions, then these more specific constructs may turn out to have greater predictive power and utility than the omnibus construct of emotional understanding (or management). Not only should emotion-specific EI scores be matched to emotion-specific criterion variables, but these may even differ across the four branches of EI (e.g., fear-management but not fear-understanding may predict job performance in jobs involving high elements of risk). Although results from the current study do not show meaningful differences in the prediction of personality or intelligence from anger-, fear- and sadness-related EI scores, this study did not include measures of well being or life success, which different emotions

might differentially predict. Certainly, results from this study indicate that it might be worthwhile to conceptualise some of the branches of EI as emotion specific, and plan the ways that these emotion-specific abilities may relate to different criteria.

4.2 Variation due to Reading Comprehension

It is clear that reading comprehension is not accounting for the specific variation shared by the STEU and STEM, in agreement with Hypothesis 2. Although the relationship of the STEU and STEM to RAPM and Knowledge decreased after reading comprehension was accounted for, this decrease was no more marked than for the correlation between RAPM and Knowledge. That is, Strategic EI is no more strongly related to reading comprehension than are other intelligence measures (reading comprehension is *not* the source of unique variation between the STEU and STEM). Chapter 5 found that cultural differences were significant before years of English speaking were accounted for, implying that some of the variation in STEU and STEM scores was the ability to read and comprehend the (English) text in the item. However, the results of the current study indicate that reading ability is not the primary source of EI-related variation. Results from Chapter 5 might indicate a threshold effect, where a certain level of reading comprehension is required before the test becomes a measure of EI rather than a measure of reading.

For both the STEU and STEM, relationships with RAPM and Knowledge declined after reading comprehension was controlled for, potentially indicating that the relationship between Strategic EI and cognitive ability is due to the reading load on the text-based Strategic EI tests. However, the STEU and STEM loaded on different factors to the cognitive ability factors in structural analysis, indicating that this source of shared variation was not sufficient to cause Strategic EI tests to assess *only* reading comprehension.

Although such tests clearly do require a threshold level of reading skill for the test-takers to adequately complete them, it is apparent from this study that EI tests measure much more than the ability to read the question. However, the fact that a threshold level of reading is required obviously limits the interpretation of Strategic EI scores for different populations (e.g., lower literacy

populations, populations from non-English speaking background). This is also an important concern when evaluating cross-cultural differences in EI. If EI is assessed with text-based performance-based tests, then what appears like a cross-cultural difference in EI may in fact be a cross-cultural difference in reading skill, particularly when one comparison group has had lower levels of formal education, or has a lower level of literacy than the other group.

4.3 Relationships of EI with Personality

Both the STEU and STEM related to Agreeableness, consistent with prior research on Strategic EI (Chapter 3; Brackett & Mayer, 2003; Brackett et al., 2004; Lopes et al., 2004; Lopes et al., 2003; Schulte et al. 2004; although not Chapter 5). However, neither measure showed the hypothesised relationship to Openness, which was found in Chapter 5 but not Chapter 3. Possibly this might indicate that the different instruments (the OCEANIC in Chapter 3 and the current chapter and an IPIP version of the NEO-PI-R in Chapter 5) are assessing slightly different aspects of the construct, emphasising the importance of considering personality correlates from multiple measures of personality (e.g., Matsumoto et al., 2000).

The ERA measures also related to Agreeableness, in contrast to Matsumoto et al. (2000), Rubin et al. (2005), and hypotheses. The RAFL but not the NIMSTIM Faces test showed the expected relationship to Openness. That is, both Strategic EI and ERA tasks related to Agreeableness but not Openness (except for the RAFL). The intelligence measures did *not* relate to Agreeableness but did relate to Openness, meaning that personality correlates differed for emotion-related and non-emotion-related and cognitive ability tasks. This supports the "different correlates" criteria for EI's distinction from Gc, outlined in Chapter 5. Neither the STEM nor the STEU related to *traits* of anxiety or stress, although Chapter 4 found that they did relate to *states* of anxiety and stress. This temporal state/trait difference may be important: states of negative emotionality may be ameliorated by intelligent manipulation or re-appraisal of the environment, whereas trait negative emotionality is by definition more stable and less amenable to manipulation. That is, relations between state anxiety

and stress and the STEU and STEM found in Chapter 4 cannot be attributed to trait anxiety and stress – the STEU and STEM are not behavioural dispositions but are predicting states of distress.

4.4 Relationships among Cognitive Tasks (Intelligence and EI)

Even though the STEM was given in ratings-based format, there was still a very high degree of relationship between STEU and STEM indicating that they are clearly measuring the same construct. Both STEU and STEM also related to the ERA measures, forming an EI factor with these that was distinct from non-emotional intelligence. It was surprising however, that RAFL and NIMSTIM faces did *not* relate to each other. Part of this may be a problem with the lack of variability in the NIMSTIM faces but, equally, emotion recognition ability may be quite distinct for different modalities. Extensions of the Gf/Gc model of intelligence to different sensory modalities support this interpretation: visual, auditory, kinaesthetic and olfactory abilities (Gv, Ga, Gk, and Go) form factors distinct from the memory or reasoning processes involved in the tasks (Danthiir et al., 2001; Horn & Stankov, 1982; Stankov & Horn, 1980; Stankov, Seizova-Cajic, & Roberts, 2001). In addition, Roberts et al. (submitted) found that a visual ERA measure (a facial expression recognition task based on the inspection time paradigm) was unrelated to the MSCEIT's Faces test. Although stimuli in the tasks were virtually identical, different abilities were required when an inspection time paradigm was used.

If the *emotional content* of ERA items predominates the processes used to perform the tasks (rather than the *sensory modality* or *method of presentation*) then an *emotion recognition ability* may be posited. However, if other factors such as sensory modality or method of presentation predominate, then tasks form part of other known abilities: the emotional content is only a surface characteristic of the task. From the evidence so far, the latter appears to be the case for the ERA measures. The auditory task (RAFL) was more strongly related to Gc (as Ga tasks generally are – Horn & Stankov [1982] found a Ga/Gc factor correlation of $r = .54$) and the inspection time task (NIMSTIM Faces) was more strongly related to Gf (as inspection time measures tend to be; see Grudnik & Kranzler, 2001) than either were to each other. However, when Strategic EI was added to

the mix, the shared variance due to emotional content was sufficient for emotional intelligence to form a factor separate to non-emotional intelligence, although factors were highly correlated (r = .73).

The strong correlation between EI and intelligence echoes the strong correlation between Strategic EI and Gc obtained in Chapter 5. With correlations in this range, the distinctiveness of EI from intelligence needs to be evaluated functionally as well as structurally. Based on Horn and Noll (1997), Chapter 5 suggested grounds for determining EI's distinction from intelligence might include age trends, neurological processes, correlates, and group differences. There is evidence for two of these conditions (different correlates and different group differences) in the current study. Personality correlates of EI and intelligence were clearly different in the current study: intelligence related to Openness but not Agreeableness, and EI related to Agreeableness but not Openness. In addition, women were superior at emotion management whereas men were superior at general knowledge (as expected from prior findings, Ackerman et al., 2001; Ciarrochi, et al, 2000; Day & Carroll, 2004; Kafetsios, 2004; Mayer, et al., 2000b).

4.5 Prediction of Academic Achievement

EI emphatically did *not* predict academic achievement beyond Gf and Gc, with the STEM not predicting any of the four measures of achievement even before intelligence or personality were controlled, and the STEU predicting UAI and psychology grades but not average mark at University (and predicting these *only* before intelligence an personality were controlled). A stronger relationship for the STEU than the STEM agreed with results from Chapter 3. Stronger prediction of psychology grade than overall grade also agreed with Chapter 3, and with Lieven et al.'s (2005) finding that situational tests predict grades in courses with interpersonal subject matter more strongly than those obtained in other types of courses.

Current results might indicate that Chapter 3's finding of incremental validity of the STEU over personality and vocabulary is probably based on the overlap between EI and Gf. That is, the STEU predicts achievement only because STEU scores overlap with Gf to some extent and Gf predicts achievement. However, the proportion of variance explained by the STEU and STEM was

far less in the current study than in Chapter 3 even *before* personality and intelligence were accounted for (4 to 6% and 1 to 4% for the STEU and STEM respectively, compared to 7 to 18% and 1 to 13% in Chapter 3). This might be partly due to restriction of range since the current sample has less variability and lower grades than the sample used in similar analyses in Chapter 3. In any case, it is reasonable to conclude that STEU and STEM scores do *not* incrementally predict academic achievement at University. However, Landy (2005) points out that academic achievement may not be an appropriate criterion for determining EI's validity, despite the large number of studies using this (e.g., Barchard, 2003; Parker et al., 2004; Petrides et al., 2004; Woitaszewski & Aalsma, 2004). A more convincing finding for the validity of EI might be Chapter 4's prediction of lesser state anxiety and stress, although intelligence was not controlled in these analyses.

4.6 Limitations of This Study

The samples used for structural equation modelling in this study were rather small for such a purpose ($N = 109$). For this reason, further research with larger samples is needed to confirm results showing the distinctiveness of EI from other mental abilities, as well as the emotion specificity of Understanding and Management. In addition, the emotion-specific models showed quite high correlations between factors, and some evidence that models overfit the data, such that accepting the structure at the one-factor level would not be unreasonable. Further clarification of this structural issue with a larger sample would be ideal.

Further limitations of this study relate to the markers of ERA and of intelligence. The NIMSTIM was an experimental measure created for the purpose of this study, and clearly suffered from ceiling effects. The NIMSTIM was developed in preference to using the existing JACBART in order to examine differences in presentation latency, as well as to include a more limited range of emotions (i.e., to exclude contempt). In retrospect, it is clear that using a well-developed existing measure such as the JACBART would have been better practice. Two of the intelligence markers (RAPM and reading comprehension) were given with time limits, meaning that speed as well as accuracy would have contributed to participants' scores. It is possible that this speed factor may have

contributed to the results of the confirmatory factor analysis (i.e., that the separation of cognitive ability from EI may in fact have been the separation of speeded tests from non-speeded tests). Replication of these results using untimed tests would be useful.

4.7 General Conclusions from Chapter 6

The conclusion from this and previous chapters seems to be that Strategic EI is distinct although strongly related to intelligence, with particularly strong relations to Gc that are partly but by no means entirely due to the text-based format of such tests. The emotional content involved in Strategic EI and ERA tasks may make tasks cohere as a unified construct but the processes involved may outweigh the content when determining whether tasks primarily assess EI as opposed to fluid or crystallized intelligence. For ERA measures, it is possible that emotional content is a surface characteristic of the test and cognitive processes involved are determined more by the mode of stimulus presentation. In addition, the emotional content of EI tests might be better conceptualised more specifically in terms of the different emotions involved in the situations. However, EI does not exhibit incremental validity in predicting academic achievement.

Chapter 7

General Discussion: Measurability, Distinctiveness, and Utility of Emotional Intelligence

Research in the preceding chapters had three main goals: (1) to determine whether standards-based scoring was possible for EI tests; (2) to examine possible method effects of response format and scoring type in EI research; and (3) to ascertain whether EI constitutes a valid new construct (i.e., that both trait and performance-based are EI distinct from existing constructs, and predict variation in emotion-related criteria not explained by existing constructs). Standards-based scoring was demonstrably possible, although results also highlight some possible limitations and important caveats that need to be considered as part of the process. Method effects of response format were also uncovered, although the strength of such effects does not appear strong enough to entirely invalidate previous results. There were no strong method effects of scoring type (i.e., consensus scoring). Lastly and most importantly, evidence of EI's validity was not overwhelming: the safest interpretation seems to be that performance-based EI constitutes new content areas within existing intelligences such as Gf and Gc (with some minor but potentially important differences from other components) and that trait EI is an expansion of personality (with particular aspects of trait EI replicating existing personality traits). There was very little evidence for the usefulness of EI as an adjunct to personality and intelligence in predicting important outcomes. These issues are discussed in detail in the paragraphs below.

1. Standards-Based Scoring in Emotional Intelligence: Possibilities and Future Directions

1.1 Standards-Based Scoring of the STEU

Using an empirically-derived structure of the emotions to score emotional understanding was quite successful as a scoring technique. Roseman's (2001) theory acted as a foundation for the STEU, resulting in a relatively reliable instrument that measured emotional understanding as well as other measures currently in use. Table 7.1 illustrates this point, comparing reliability estimates for the STEU obtained in this thesis to those of the two Understanding tests from the MSCEIT. Reliability is approximately equal, although the STEU can be scored according to standard rules about emotions whereas the MSCEIT tests cannot. This standards-based scoring is a strong advantage in that the reason a particular option is correct can be unambiguously and precisely defined. This allows a clear verbal specification of what construct test scores actually capture, so that EI may meet Mayer et al.'s (2000b) conceptual criteria for an intelligence.

The STEU related to emotion related criteria such as Emotion Perception, Emotion Understanding, Emotion Management, emotionally-oriented thinking, and both the physical symptoms and subjective experience of state anxiety. However, scores were not related to state depression and stress, nor to aspects of alexithymia other than emotionally-oriented thinking. In addition, very strong relationships with Gc might indicate that the STEU (and Understanding generally) forms a primary mental ability of Gc rather than a distinctly different construct.

Table 7.1

Comparison of Reliability Estimates for Understanding and Management Tests: MSCEIT versus STEU and STEM

	Mayer et al. (2003)		*Roberts et al. (submitted)*	*Barchard (2003)*	
MSCEIT	*Consensus*	*Expert*	*Consensus*	*Consensus*	
Branch 3: Understanding					
Changes	.70	.68	.52	.52	
Blends	.66	.62	.50	.54	
Branch 4: Management					
Management	.69	.64	.52	.70	
Relations	.67	.64	.51	.79	
STEU/STEM	*Chapter 3*	*Chapter 4*	*Chapter 5 (alpha)*	*Chapter 5 (test-retest)*	*Chapter 6*
	Expert	*Consensus*	*Consensus*	*Consensus*	*Expert*
Branch 3: Understanding					
STEU	.71[a]	.51[a]	.55	.66	.70[a]
Branch 4: Management					
STEM (MC)	.68 - .74[b]	.64	.50	.55	
STEM (rate)	.86 - .92[b]				.74

[a]STEU scored by standards derived from appraisal theory (not by expert judgment). [b]Multiple possible expert scores were calculated for these tests.

1.2 Standards-Based Scoring of the STEM

Although the STEM could be scored relatively reliably under expert or consensus rubrics (see Table 7.1) attempts at standards-based scoring were not successful, with reliability estimates at virtually zero. The scoring key for the STEM consisted of a rank order list of coping strategies for five types of situations, developed from empirical studies on the effectiveness of different types of coping. There are at least three types of qualitative judgments involved in this approach, and each might introduce a source of human error: (1) judging the equivalence of concepts outlined in different

scientific models of coping; (2) matching the STEM items to the five types of situations; and (3) matching the STEM response options to coping strategies in the scoring key (such that rank ordering their effectiveness is possible).

Both situation types and types of coping strategies from different researchers were often based on different models, theories or assumptions, necessitating qualitative judgments of their equivalence. If such approaches are taken in future, restricting empirical evidence to *one* paradigm may reduce or eliminate this source of error (e.g., studies only using the COPE, Carver et al., 1989; or only using the same version of the Ways of Coping Checklist; Folkman & Lazarus, 1980, 1985). Qualitative judgments were also necessary for categorising the STEM's items as representative of a particular situation, and the STEM's response options as representative of a particular type of coping.

It may be that a theory-driven scoring key and an empirically driven test construction methodology (the SJT) are simply not compatible. To use a theory-driven scoring key, each item must clearly represent a situation type within the scoring key and each response option a response type from the scoring key. When both items and response options are generated by free responses from a large number of people, erratic or idiosyncratic details in both item and response options are created that makes it difficult to match these back to pre-determined "types". One possibility to make the SJT test construction methodology more compatible with a theory-driven scoring key is to make the test construction process more theory driven at the outset. The interview process for item generation could be strictly modelled on the types of situations used in the scoring key. Similarly, response options could be generated according to *types* of responses represented in the scoring key.

A more serious concern than standards-based scoring being incompatible with the SJT approach to test construction is the difficulty of justifying standards-based scoring for the Management construct, where there simply may not be definitively applicable standards. It is possible that any discrimination of situation and response types broad enough to allow general rules about response effectiveness are too broad to be applied to specific situations. Performance-based emotion management items are currently all presented as quite specific hypothetical situations with several

possible responses (Freudenthaler & Neubauer, 2005; Mayer et al., 1997; Mayer et al., 2002b). Emotion management might also be measured by knowledge of the general rules of what type of response is best in what type of situation, rather than the application of these general rules to specific instances (i.e., situations would be presented as a general type, rather than a specific example). Both MEIS and MSCEIT contain these "general type" items in the Understanding branch, as does the STEU in the form of its abstract or context-less items. This specificity-of-context distinction between Management and Understanding constitutes another difference between Management and Understanding scores that may be methodological rather than construct-related and as such problematic.

Table 7.2 outlines the empirical differences between Understanding and Management; the methodological differences in their measurement; and the conceptual differences in the processes involved in responding to Management and Understanding items. Such differences are important in considering whether differences are construct or method-related, and in discussing the meaning of results found.

Table 7.2

Empirical, Conceptual, and Methodological Differences between Understanding and Management of Emotion

Understanding	*Management*
Empirical Differences	
➢ More strongly related to intelligence (is probably a PMA of Gc) than Management is ➢ Small to zero gender differences ➢ Predicts state anxiety (but not depression or stress)	➢ More strongly related to Agreeableness (personality) than Understanding is ➢ Marked gender differences favouring females ➢ Predicts state anxiety, depression and stress
Conceptual Differences	
➢ Task is to interpret what happened ➢ Oriented exclusively to the past (events have already happened, one's task is to interpret them) ➢ Judgments involved : (1) interpreting appraisal dimensions (e.g., interpret events as representing high/low control potential, motive consistency etc)	➢ Task is to interpret what *should* happen ➢ Both past and future orientations (events have already happened, but responses to these events will happen in the future) ➢ Judgments involved: (1) personal goals/values determine what constitutes the effective handling of an emotional situation; (2) interpreting appraisal dimensions (e.g., lack of social contact must be interpreted as motive inconsistent)
Methodological Differences	
➢ Measured exclusively with multiple-choice tests ➢ Standards-based scoring possible ➢ Can be measured by items with broad general content as well as specific situations	➢ Measured exclusively with ratings-based tests in the MSCEIT, or with both ratings-based or multiple-choice tests in the STEM and MEIS ➢ Standards-based scoring not possible (at present) ➢ Cannot be measured by items with broad general content, but only by items describing specific situations (at present)

It is possible to create management items that resemble a "general type of situation", rather than the specific situations currently used as items. An example item that contains general, broad content rather than specific content might resemble the following. *When dealing with a recurring or chronic illness, the best thing to do is: (a) get support from friends or relatives, (b) seek information so as to be well-informed, (c) avoid thinking about it, (d) think or imagine things are better.* This

comes straight from the first situation in the scoring key outlined in Study 4 of Chapter 2. Options (a) and (b) are seeking social support and problem-focused (information seeking) and are therefore superior to options (c) and (d) which are avoidance and wishful thinking. This type of item may be more amenable to standards-based scoring derived from empirical research, and might also reduce the methodological differences between Understanding (which is measured with both broad-content and specific-content items in the both the STEU and MSCEIT) and Management (which is measured with specific-content items only in both the STEM and the MSCEIT).

1.3 General Conclusions about Standards-Based Scoring

Examination of the STEU showed that using empirical evidence as the foundation for standards-based scoring is both possible and useful in situations where detailed enough evidence exists. However, empirical evidence as a standards-based scoring approach may be fundamentally incompatible with SJT approaches to test construction, or tests where items are particularly detailed, idiosyncratic, or context-bound. The much greater difficulty in standards-scoring the STEM than the STEU highlights the importance of using only one research paradigm (rather than several different traditions) in creating evidence-based methods of scoring.

2. Consensus Scoring and Ratings-based Response Formats as "Method Effects"

To reiterate, virtually all evidence on performance-based EI has been based on the MSCEIT or its earlier manifestation, the MEIS (Mayer et al., 1997, 2002b). Both these batteries have both an idiosyncratic scoring type (consensus based scoring, using either normative or expert samples) and an unusual response format for most tests (ratings-based scales for all tests bar two assessing Understanding). This thesis examined consensus scoring in some detail, finding that gender differences were *not* due to the number of men or women in the screening sample, and that consensus scoring was *not* acting as a strong method effect. That is, consensus scoring does not appear problematic as a method effect or as a cause of group differences. The ratings-based response format was also put under the microscope. Individual differences in the use of the ratings-scale were found, as was evidence that the ratings-based format acts as a method effect in distinguishing Strategic EI

from intelligence, and Understanding from Management. Implications of these findings and suggested solutions to the response format issues are discussed below.

2.1 Validity of Consensus Weightings: Group Differences?

This thesis examined whether the gender composition of the screening sample used to derive consensus weights may result in gender differences when these weights are applied. Two studies in this thesis compared gender differences on the STEM under two different sets of scoring weights derived from samples with different gender compositions (Study 3 of Chapter 2 [see Part 2, Section 3 of Chapter 2]; and the study outlined in Chapter 5). Chapter 2 found that the female superiority on the STEM was no different when scored with female only expert weights than male only expert weights (non-significant in both cases). Results reported in Chapter 5 found a significant female superiority on the consensus scored STEM even when consensus weights were derived from an entirely male sample. This is a re-assuring finding for one aspect of the validity of MSCEIT scores: Gender differences do not appear to be due to gender differences in the screening sample used to determine scoring weights. Female superiority found on the Management branch of the MSCEIT does then seem to be a genuine finding rather than a methodological mirage (e.g., Day & Carroll, 2004; Kafetsios, 2004). Possible reasons for these gender differences are discussed in Section 2.5 of this chapter and the ramifications and practical applications are discussed in Section 3.1.3.

2.2 Consensus Scoring and the Correlational Criteria for an Intelligence

The correlational criteria states that EI should be "similar to, but distinct from, mental abilities described by already established intelligences" in order to qualify as an intelligence (Mayer et al., 2001, p. 270). Davies et al. (1998) suggested that the distinction of earlier Perception measures of EI from intelligence was due to consensus scoring rather than the construct measured (i.e., that an EI factor separable from other intelligence factors was really a consensus-scoring method factor). This assertion was tested empirically in Chapter 5, where EI and intelligence constituted separate factors even when all tests were consensus scored, meaning that the EI's distinctiveness could not be due to consensus scoring. However, the correlation between EI and Gc factors was slightly higher

when Gc as well as EI was consensus scored (r = .79 compared to r = .73). This suggests that consensus scoring may be acting as a method effect, where consensus-scored tests correlate more highly with other consensus-scored tests irrespective of the construct measured, although such an effect is quite small.

Given that consensus scoring does not constitute a major method effect and given the difficulty of scoring emotion management tests according to theoretical or empirical standards, consensus scoring may be a reasonable approach to scoring such tests, if measurement problems such as distribution shape could be ameliorated (MacCann et al., 2003; 2004). One alternative may be to generate consensus weights from non-proportional algorithms. For example, mean ratings of each option might be used to score multiple-choice tests, rather than the current practice of using the proportion of the people selecting each option (note that this involves the screening sample completing the test in ratings-based form, but the test-takers completing the test as multiple-choice; see Table 3.3, Chapter 3). In the MEIS and MSCEIT the current procedure is for the screening sample to select the best response, and option weights correspond to the proportion selecting each option (e.g., If 82% of the screening sample select (*a*), (*a*) is awarded as score of 0.82). Instead, the screening sample could *rate* each response such that the mean of their ratings could form the option weight for that response option (e.g., If the screening sample awards (*a*) an average rating of 7.5/10 then (*a*) is awarded a score of .75). Although such methods are still blindly atheoretical, the proportion selecting an option is less strongly related to the option weight than in proportional scoring, which may help the distributional properties of the test scores.

2.3 Validity of Ratings-Based Responses: Individual Differences in Rating-Scale Use?

Individual differences in the use of the rating-scales were examined, ascertaining whether bias towards the high or low end of the scale, or a tendency to spread out one's responses along the scale related to individual differences in known personality traits. Chapter 3 found that higher raw ratings on the STEM were given by non-experts (compared to experts), and females (compared to than males). Higher raw responses on both the STEM and Stories test also related to Agreeableness. There

are systematic differences in the way individuals use the rating scale and some of the reasons for these differences may depend on the nature of the stimuli.

The link between a high ratings and the personality trait Agreeableness may be an indication of acquiescence, where participants have a response bias towards agreeing with the statement. This has long been known to be a problem in personality testing, particularly in separating cultural differences in personality from possible cultural differences in use of the rating scale (Couch & Keniston, 1960; van Herk, Poortinga, & Verhallen, 2004). The usual solution is to include equal numbers of positively and negatively keyed statements to balance any overly defensive or suggestible responding. However, acquiescence may also be a problem in *ability* tests that use rating scales, and the solution is not quite as simple. A solution suggested and undertaken in this thesis was the use of profile scoring to control for individual differences in the use of the rating scale. In addition, rating scales used in EI tests were anchored by a descriptor at each scale point in this thesis, which should further minimise individual differences in rating-scale use.

These possible individual differences in rating scale use have important implications for the MSCEIT, where profile scoring is not used, and where scales are anchored in an inconsistent way. In the Perception branch, the Faces test is anchored at each end, whereas the pictures test is anchored at each point with a different emoticon. Both tests from the Facilitation branch are anchored at each end, and both tests in the Management Branch are anchored at each scale point. The possibility of individual rating scale use contaminating scores on the MSCEIT may be particularly important in cross-cultural research using the MSCEIT, as differences in rating-scale use have been found to differ systematically across cultures (van Herk et al., 2004). Indeed, this may be a problem in examining any form of group differences where the groups might conceivably use the rating scales differently. For the MSCEIT, this might be particularly problematic for gender differences on the Management tests given that: (a) there are gender differences in rating scale use for emotion management items (see Chapter 3), and (b) no adjustment is made for possible systematic differences in rating-scale use when calculating MSCEIT scores. However, gender differences on the ratings-based STEM (where

adjustments were made) were similar to gender differences on MSCEIT Management, indicating that such differences are not due solely to differences in rating-scale use. Nevertheless, this remains an important concern in interpreting the magnitude of gender differences as well as determining differences between other groups (e.g., age differences, cultural differences).

2.4 Response-Format and the Correlational Criteria for an Intelligence

Just as Davies et al. (1998) suggested that consensus scoring may be the real source of variance separating EI tests from other cognitive abilities, it may be that the ratings-based response format of most of the MSCEIT's tests is the real source of variance distinguishing between EI and other cognitive abilities. The fact that Understanding tests have the strongest degree of relationship to intelligence, and are the only tests in the MSCEIT to be multiple-choice rather than ratings-based adds weight to this possibility.

This thesis found some support for this proposition. Results from Chapter 3 indicated that the multiple-choice STEM correlated more highly with the (multiple-choice) vocabulary measure than did the ratings-based STEM (r = .26 compared to r = .40; however, this difference was not significant). That is, identical test content related more strongly to intelligence when presented in multiple-choice format. However, separate factors for EI, Gf, and Gc were found when all EI indicators were multiple-choice, such that the distinction between intelligence and Strategic EI could not be due to the difference in response format alone (see Chapter 5). Nevertheless, the correlation between EI and Gc factors was very high in this situation (r = .73), indicating considerable overlap of the constructs. A sensible interpretation would be that Strategic EI is a new primary mental ability of Gc. One possible caveat to this conclusion is that the Strategic EI factor in Chapter 5 was predominantly defined by Understanding rather than Management (two of the three indicators were Understanding and they loaded more highly than the Management indicators on the EI factor). Thus, Understanding might best be interpreted as a PMA of Gc, although Management might be something different.

Thus, there is equivocal support for the idea that ratings-based scales may be partially responsible for the distinction between EI and intelligence. In conjunction with evidence of individual differences in rating scale use, this makes the use of rating-scales to assess EI (or other constructs) appear problematic. On the other hand, rating scales do seem to be considerably more reliable than multiple-choice formats, as Table 7.1 illustrates.

Further adding to the difficulty in deciding which response format is more appropriate is the possibility that the response format chosen may change the nature of the construct measured. Selecting the best response to an emotionally-laden situation may be conceptually different from judging a variety of responses. In Management tests, multiple-choice items require knowledge of the most effective response in a given situation, whereas ratings-based items require differentiation between effective responses, mildly effective responses, benign but ineffective responses, and highly destructive ineffective responses. These may be different processes, and could conceivably predict different things. Not knowing how to avoid horribly destructive behaviours might feasibly lead to more problems in one's life than not knowing what the best behaviours are. Conversely, knowing the best things to do might lead to more success than merely knowing how to avoid the worst responses. An interesting future direction for methodological research would be to assess if success-related criteria were better predicted by multiple-choice tests, and (lack of) failure-related criteria were better predicted by ratings-based scales. Such methodological research goes beyond the construct of EI, and could fruitfully be applied to SJTs in other areas, as well as social and practical intelligence.

2.5 Response-Format and the Distinction between Understanding and Management

One of the major aims of the current thesis was to determine whether the empirical differences between Understanding and Management were construct-related or method related, and one of the primary candidates for a method effect was the difference in response format. In Chapters 3 and 4, Understanding and Management were much more strongly related when both were administered in multiple-choice format than when Management was in ratings-based format. Indeed, multiple-choice Management and Understanding were almost indistinguishable in Chapter 3.

However, there are some clear empirical differences between Understanding and Management outlined in Table 7.2 that indicate these are *not* the same construct: (1) Understanding related more strongly to intelligence (even when controlling for response format); (2) Management relates more strongly to Agreeableness; (3) Female superiority is much more pronounced for Management than Understanding; and (4) Management and Understanding differentially predicted states of anxiety, depression, and stress.

2.5.1 Intelligence Relates More Strongly to Understanding than to Management

The first empirical difference between Understanding and Management is Understanding's stronger relationship to intelligence. Chapter 3 found this greater relationship only when Understanding was measured with multiple-choice items and Management with ratings-based items (i.e., *not* when both STEU and STEM were multiple-choice). That is, Understanding's greater relationship to intelligence was at least partly due to response format. However, Chapter 5 found Understanding related more strongly to intelligence than Management, even though all tests were multiple-choice. Taken as a whole, these results indicate that Understanding *does* relate more strongly to intelligence, irrespective of response format issues, but that the magnitude of this difference may be exaggerated when Management is measured with ratings-based tests. Given that this is the case for the MSCEIT (Management tests are all ratings-based), many past findings of this stronger relationship may be exaggerated. An interesting future direction would be to administer the MSCEIT Management tests in multiple-choice format and compare findings with the current ratings-based tests.

2.5.2 Agreeableness Relates More Strongly to Management than Understanding

A second difference between Understanding and Management is Management's stronger relationship to Agreeableness. Relationships with Agreeableness were stronger for the STEM than the STEU in both Chapter 3 and in Chapter 6, across both ratings based and multiple-choice versions of the STEM. Although these differences were not significant, differences in correlation require a very large effect size to ensure statistical significance, and the correlation between STEM and

Agreeableness is small to begin with. Management's stronger relationship with Agreeableness is consistent with much prior research using the MSCEIT, and as such is likely to be a genuine finding rather than chance variation (Brackett & Mayer, 2003; Brackett et al., 2004; Lopes et al., 2003; 2004; Schulte et al. 2004). Given that a key conceptual difference between Understanding and Management is the influence of personal goals and values, this may be where Agreeableness operates to influence scores on Management more so than Understanding tests. Chapters 5 and 6 found this Agreeableness relationship particularly for the dimension Morality (equivalent to the NEO-PI-R Straight-Forwardness), a dimension concerning whether people use others for their own ends. This source of individual differences in values may well influence scores on Management tests through inference of goals and values in emotional situations. Such an interpretation agrees with Mayer et al. (2001) supposition that both Understanding and Management are cognitive abilities but that Management must also "balance many factors including the motivational, emotion and cognitive" and has an "interface with personality and personal goals" (p. 235).

A second reason why Management may relate more strongly to Agreeableness than Understanding might relate to an investment theory of positive temperament into EI. Zeidner, Matthews, Roberts, and MacCann (2003) suggest positive temperament facilitates rule-based learning of emotional heuristics and rules in children. Similar processes may occur for adults. For example, Agreeable people may be both more motivated to ameliorate negative emotional situations (and hence develop knowledge of emotion management) and have more opportunities for learning through their interest in others' emotions (i.e., the sympathy component of Agreeableness). Positive temperament may also facilitate or motivate adults to invest their emotional knowledge and understanding in developing strategies for managing emotional situations (i.e., more Agreeable people use their Understanding to develop their Management). The greater differentiation of Understanding and Management at higher ages (found in Chapter 4) supports an interpretation that Understanding may be invested to increase Management.

An awareness of why Understanding relates more strongly to intelligence but Management relates more strongly to personality can be gleaned from examining the judgments required by Understanding and Management items, outlined in Table 7.2. Both Understanding and Management items require test-takers to process situational details as representative of particular appraisal dimensions and therefore of particular emotions, which may involve elements of personality or ways of viewing the world. For example, one person may view playing the stock market as involving relatively high control potential and someone else may view the same situation as having little or no control potential. The standards-based scoring system used in the STEU implicitly assumes that there is an ideal or intelligent way to interpret particular events (e.g., there is a correct answer to how much control potential there is in playing the stock market!), and that emotional knowledge rests on this a true state of affairs. To the extent that this is true, the STEU thus primarily assesses knowledge (and the fact that the STEU correlates so strongly with Gc tasks attests that there is an ideal interpretation of emotional events that represents knowledge).

The STEM also involves appraisal of events as representing particular emotions to a certain extent – it would help to know whether you are sad or angry in order to know how to effectively ameliorate this unpleasant feeling. However, the STEM (and other Management tests) also involves value judgments about the ideal state of being, equilibrium, or final goal. These judgments are clearly non-cognitive to some extent. For example, whether one cares if other people experience negative emotions is a value judgment that may influence one's answers. A person very low on Agreeableness may not consider a crying or angry workmate as disturbing their equilibrium, whereas someone high on Agreeableness might view this as a serious disturbance. The assumption that Management items can be scored as correct or incorrect rests on the assumption that one particular set of values is more adaptive, useful or intelligent than others. Mayer et al.'s (2001) twin claims that Management is an intelligence and that Management involves elements of personality assume at some level that aspects of personality feed into more intelligent behaviour. The concept of emotion management thus rests on the middle ground between personality and intelligence.

2.5.3 Gender differences are Larger for Management than Understanding

A third difference between Understanding and Management was the magnitude of gender differences. Chapters 5 and 6 found significant female superiority for Management but not Understanding (which remained when controlling for the proportion of females in the screening sample). Given that the strongest gender difference occurred when Understanding and Management were measured in multiple-choice format (in Chapter 5), this gender difference is clearly not due to differences in response format. There are several possible causes of such gender differences: (1) the greater proportion of context-bound items advantaging women due to the female advantage in contextual memories; (2) gender differences in Management are due to gender differences in Agreeableness, which Management tests overlap with; and (3) that the test content is biased as the sample that generated it was mostly women.

(1) Gender differences in Management are due to episodic memory. Greater gender differences in Management than Understanding may relate to the situational/contextual quality of Management items compared to Understanding items. Both the STEM and the MEIS and MSCEIT Management tests include only context-bound items (i.e., items consist of a situation with a protagonist doing something). In contrast, the STEU and Understanding tests from the MEIS and MSCEIT include items that are not context bound but ask about abstract, factual information about emotions (e.g., *By their own actions, a person reaches a goal they wanted to reach. How is the person most likely to feel?*). There is some evidence that women are superior to men on tasks involving episodic and autobiographical memory, particularly when emotion is involved, and this superiority may be the reason women out-perform men on Management but not Understanding tests (Davis, 1999; Ross & Holmberg, 1992; Seidlitz & Diener, 1998).

(2) Gender differences in Management are due to Agreeableness. If Agreeableness or Morality does impact on STEM scores through the inferred values of the items' protagonists, then gender differences in Management may be due to gender differences in Morality. However, empirical

results indicate that this is not the case. Both in Chapter 5 and in Chapter 6, female superiority remained when accounting for Morality. Thus, this does not seem to be the most likely explanation.

(3) Gender differences in Management are due to the mainly female sample that generated item content. STEM items were generated by a sample that was 67% female and STEM response options by a three samples that were 47%, 53%, and 71% female, respectively (see Chapter 2). Although the gender of the items' protagonists was controlled so that half were male and half female, it is possible that item content may have been more "female friendly" since more of it was created by females than males. The major argument against this is that the STEU surface content was created entirely by one female (the author) and yet showed no such gender difference. In addition, the item content of the MSCEIT's Management tests was presumably created by the three male test authors and yet a female advantage is still found for these tests.

It seems the most likely explanation for these gender differences is the effect of context advantaging women through their greater memory for emotional events. Implications and practical application of these gender differences are discussed further in Section 3.1.2.

2.5.4 Understanding and Management Predict Different Aspects of State Distress

Understanding and Management also differentially predicted different states of distress. Understanding predicted lower state anxiety (particularly the subjective feelings of anxiety) whereas Management also predicted the amelioration of states of depression and stress. Such a relationship makes sense when considering the likely consequences of being able to understand and manage emotional situations, respectively. If one understands emotional events, there is less uncertainty in the interpersonal domain about what will happen next. If one knows effective management strategies one will better be able to control emotional situations. Since anxiety is caused by uncertainty of future events and depression and stress contain elements of low control potential high Understanding should lower anxiety and Management should lower depression and stress.

It seems that Management and Understanding are different kettles of fish, despite the strong relationship between them. Evidence so far indicates that Understanding is probably a PMA of Gc,

relating strongly to Gc, showing similar relationships to personality as Gc does (relating to Openness more so than Agreeableness), and showing similar gender differences to Gc. Although Management also relates to intelligence (albeit less strongly than Understanding), there are also relationships with Agreeableness, and a large female superiority. Management differs more than Understanding from conventional intelligence. The inclusion not just of processing of emotional information but using value judgments to act on this information may be the important difference.

3. Validity of Performance-Based EI

3.1 Is Performance-Based EI Distinct from Intelligence?

Although Strategic EI is clearly strongly related to cognitive ability, and particularly Gc, relationships do not appear to be due only to the verbal text-based format of the tests. Results from Chapters 3 and 6 demonstrated that the relationship of EI to intelligence is not simply due to the verbal ability involved in answering EI questions. Chapter 3 found that relationships among EI were not due to vocabulary-related variation alone, and Chapter 6 found that relationships among EI tests and between EI and intelligence were not due only to reading comprehension ability. In addition, the STEU was much more strongly relate to Gc than the STEM, and the STEM had a heavier load for reading, containing more text and longer response options.

Nevertheless, Strategic EI is clearly strongly related to cognitive ability and particularly Gc. In Chapter 5, Strategic EI and Gc factors were very strongly related and in Chapter 6 broad EI and intelligence factors were also strongly related. Given factorial independence but a very strong relationship between EI and intelligence, it was suggested in Chapter 5 that EI's distinction from intelligence should be evaluated on functional as well as structural grounds, based on Horn and Noll (1997). Information gained in this thesis addresses to developmental grounds, distinct correlates grounds, and group differences. Many of the issues pertinent to whether EI is distinct from intelligence are the same ones discussed earlier in terms of whether Understanding and Management are different constructs. Collectively, these findings indicate that Management but not Understanding might be distinct from Gc.

3.1.1 Developmental and Age Trends for EI

A comparison of Chapter 3's sample (median age 19) with Chapter 4's sample (median age 33) shows that both STEU and STEM scores seemed to increase with age from young to middle adulthood, as Gc does (Horn & Cattell, 1967). However, Chapter 5's third year university sample did even better than the adult volunteers used in Chapter 4, making it difficult to clarify any age trends. A tentative conclusion might be that EI increases over the lifespan, but samples of different age groups were not random (first year psychology students, volunteers motivated by a desire for feedback on their emotional intelligence, and third year psychology students completing an optional course on psychological assessment). Clearly, ability differentiation within these three groups might well be due to sampling issues in addition to age differences.

3.1.2 Personality Correlates of EI

Chapter 5 also found a Strategic EI factor related to Openness to Experience in much the same manner as Gc does, correlating with the intellectual/imaginative facets rather than the emotional/creative aspects. However, this factor was primarily defined by Understanding tasks rather than Management tasks. In Chapters 3 and 6, both STEU and STEM scores also related non-trivially to Agreeableness, which Gc does not (Ackerman & Heggestad, 1997). Relationships to Agreeableness were stronger for the STEM than the STEU in both instances. On these grounds, there does appear to be a difference between Management and other cognitive abilities, but perhaps not between Understanding and such abilities.

3.1.3 Gender Differences in EI

Gender differences on the STEM were consistently in the opposite direction to differences on Gc tasks (female superiority for Management, male superiority for Gc). This female superiority was also found for the STEU, Stories, and ERA tasks but only to a much smaller extent, where differences did not reach significance (and in fact a small non-significant male superiority was found for the MEIS Understanding tests). This female superiority might be potentially important in applied settings. Chapter 5 suggested that the advantage of EI may be in its consequential validity (see

Messick, 1995) rather than in incremental prediction. For example, Chapters 3 and 6 of this thesis indicate that EI does predict some variation in academic achievement (although not incrementally over existing intelligence measures) and could conceivably be used as part of the selection process for academic programs. Males currently score more highly than females on the Scholastic Aptitude Test (SAT; Allspach & Breining, 2005), as well as standard tests of reasoning ability which involve elements of spatial ability (e.g., Raven's Progressive Matrices, Colom et al., 2004). Alternative assessments that reduce this adverse impact for females might thus be a useful application for Management tests.

3.2 Do Abilities from Different Branches of EI Form the One Construct?

Evidence from Chapter 6 provides weak evidence that nonverbal measures of emotion recognition and verbal measures of Strategic EI do form a single "EI" construct. Confirmatory factor analysis supported a model where EI consisted of both ERA (Emotion Recognition Ability) and Strategic EI and was separate from cognitive ability. However, the auditory ERA measure appeared to relate more to Gc and the inspection time measure more to Gf than either did to each other (in agreement with Roberts et al., in press). That is, although the emotional content of the Strategic EI and ERA measures was sufficient to make them cohere as a single factor within the structure of mental abilities, the processes involved (particularly for the ERA measures) might equally make such tasks measures of Gf or Gc.

The real question here is whether emotional content is salient enough to make it a defining characteristic of what the tasks measure or merely a surface characteristic (with the processes involved defining what the tasks measure). Data from this thesis indicates that emotional content of tasks is probably somewhere between a defining characteristic and a surface characteristic of tasks. Although knowledge-based EI tasks (Strategic EI) related strongly to acculturated knowledge (Gc) and ERA tasks based on information processing related most strongly to other information processing abilities (Gc for the RAFL and Gf for the facial recognition task), these tasks did hang together independently from other cognitive abilities. There is some evidence that tasks with emotional

content form a distinct area within the structure of mental abilities, but equally, there is evidence that the processes involved in these tasks are the same as those involved in existing abilities. Faceted models of intelligence such as the Berlin Intelligence Structure (BIS) may be useful in interpreting these process-content distinctions (Beauducel & Kersting, 2002). In this model, both process and content are considered "real" sources of variation in determining what tasks measures, such that both emotional content and the differing processes of the ERA tasks (the inspection time paradigm for Faces or auditory processing for the RAFL) can simultaneously be considered defining characteristics of the task. Viewing results through the lens of the BIS model, it might be concluded that emotional content forms a new area to the verbal, figural and numerical content already included in the BIS Model.

3.3 Convergent and Criterion Validity Evidence for Emotional Intelligence

As Landy (in press) points out, the distinctiveness of EI from existing constructs does not indicate the validity of EI without corresponding convergent or predictive validity evidence. The concentration on whether EI meets the correlational criteria for an intelligence, while important, has somewhat obscured the point that EI must also uniquely and meaningfully relate to appropriate emotion-related variables in order to be a useful construct. Generally, there were few relationships with well-being measures and those that were there were of small magnitude. Also, the more directly the criteria related to emotions and emotional states, the more strongly the STEU and STEM predicted them. The STEU and STEM predicted state anxiety, depression and stress as well as externally-oriented thinking, but did not consistently predict life satisfaction, loneliness, mental health or social and health-related lifestyle variables. That is, EI does not predict a broad range of positive outcomes but may be useful for predicting specific, carefully chosen emotion-related criteria.

3.3.1 Relationships with State Anxiety, Depression, and Stress

A key piece of evidence that EI does meaningfully predict emotion-related criteria is the STEU and STEM's relationship to states of anxiety, depression, and stress. Chapters 5 and 6 found no relationship between the STEU or STEM and *traits* of anxiety, depression, or stress, but Chapter 4

found that the STEU predicted *state* anxiety reasonably strongly, and the STEM predicted state anxiety, depression, and stress. That is, relationships to state distress cannot be due to the STEU or STEM's overlap with personality, but rather are due to the ability of people to understand emotional events that reduces states of anxiety, and the ability of people to use effective management strategies to reduce anxiety, depression, and stress. However, intelligence and personality were not controlled in Chapter 4, so it is possible that this prediction may be an effect of intelligence rather than EI.

3.3.2 Relationships with Alexithymia

Both the STEU and STEM related to the externally-oriented thinking component of alexithymia, but the relationship remained after intelligence and personality were controlled in only one of the two studies examined (in Chapter 3 but not in Chapter 5 where short forms of the STEU and STEM were used). There were no consistent relationships with the other aspects of alexithymia (difficulty identifying and describing feelings). This result contrasts with prior findings that only the difficulty identifying feelings component of alexithymia relates to performance on an emotion recognition task involving the recognition of facial expressions of negative emotions (Parker, Prkachin, & Prkachin, 2005). It seems that different components within EI predict different components of alexithymia (i.e., Emotion recognition relates to less difficulty describing emotions, whereas Strategic EI relates to a more concrete non-emotional thinking style).

These different components of EI require different levels of cognitive processing: perception/recognition of emotions requires lower level visual processing of information, whereas Strategic EI requires more cognitively complex manipulation of the stimuli (Mayer et al., 2001; Roberts et al., 2001). As such, it makes sense that perception/recognition abilities predict more basic deficits in emotion processing (difficulty in describing one's feelings) whereas strategic abilities predict a cognitive style that discounts or ignores emotional information. Again, these results illustrate the point that criteria may need to be stipulated at quite a specific level. Both EI and alexithymia consist of several different underlying components. A one-to-one specification of

predictor and criteria (i.e., "EI predicts alexithymia") may be too simplistic to be meaningful or measurable.

3.3.2 Relationships with Other Well-Being Indicators

The other well-being indicators examined in this thesis are less closely aligned with emotions and emotionality and there were no strong or consistent relationships with these variables: life satisfaction, loneliness, mental health, and selected life-style variables (income, job status, amount of social contact, and smoking and drinking behaviours). There may have been problems with the General Health Questionnaire used in Chapter 3, which measured mental health (the sample scored unusually highly on this measure). There may also have been range restriction issues in collecting life-style information from the community sample in Chapter 4, who had very high socio-economic status (based on level of education and wages) and were of necessity all computer literate. However, in conjunction with the lack of relationship to other variables, a sensible conclusion would be that there are only small or inconsistent relationships between these well-being measures and EI. The criterion validity of performance-based EI may be limited to a small range of outcomes directly related to emotions and emotional states.

3.3.4 Relationships with Academic Achievement

Although Chapter 3 found that the STEU predicted up to 7% of the variation in academic achievement after vocabulary and personality were controlled, this finding was not replicated in the study reported in Chapter 6. Relationships between EI and academic achievement were much weaker in this study than in Chapter 3, even before accounting for personality and intelligence. Although EI does relate to academic achievement, scores do not incrementally predict achievement after Gf and Gc are adequately accounted for. EI's prediction of academic achievement seems to be due to the shared variation between EI and intelligence, rather than EI's relationship to adjustment the emotional stressors of academic life.

Certainly it seems that Strategic EI's power to predict criteria is restricted to modest relationships to variables directly related to emotional states and behaviours (i.e., emotionally-

oriented thinking, and state anxiety, depression, and stress). Broader indicators of positive social and successful functioning such as academic achievement and life satisfaction are not well predicted by EI. It may be that tests are simply too similar to other forms of intelligence (particularly Gc and Understanding) to meaningfully predict criteria once intelligence is accounted for.

4. Validity of Trait EI

Chapter 5 of this thesis found that the major component of commonly used measure of trait EI (Schutte et al.'s SSRI, 1998) was a replication of Extraversion, although minor components did appear relatively distinct from known personality traits. Although the SSRI had some major overlap with existing personality traits, results support the possibility that trait EI may tread new ground in the personality field. Similarly to the performance-based Strategic EI measures, the SSRI predicted the externally-oriented thinking component of alexithymia, but in contrast to performance-based EI, this relationship was still significant after controlling for personality and intelligence. In addition, the Social Skills factor incrementally predicted the two other components of alexithymia (difficulty identifying and describing feelings) as well as state loneliness. That the Social Skills component was the most predictive is somewhat troublesome, as this factor was not part of the stated content domain that the instrument set out to measure – Social Skills was later uncovered via structural analysis, with some disagreement over whether this exists at all (Austin et al., 2004; Gignac et al., 2005; Petrides & Furnham, 2000; Schutte et al., 1998). Nevertheless, the potential usefulness of such a construct might suggest a careful theoretical analysis and re-design of the instrument to measure this factor more clearly.

Despite the large amount of research being generated in the trait EI area, the label "emotional intelligence" is particularly misleading here. Trait EI measures are emphatically not cognitive abilities but an expansion of the personality field into emotion-related dispositions (other than the emotional lability currently covered by Neuroticism). Although such measures may reliably predict emotionally intelligent behaviour they might better be referred to as emotional sensibility or a similar term, in the interests of accuracy.

5. Alternative Directions for the SJT Instrument Construction

In the current thesis, a deliberate policy of carefully limiting and defining the Management construct within narrow parameters guided the SJT test construction process. On the one hand this made for clearer definition of the construct, but on the other hand the lack of subjectivity and idiosyncratic detail in the situations that formed items meant a loss of ecological validity. In constructing the STEM, item length was limited to two sentences (to ensure items did not assess reading ability), emotional content was limited to three emotions (which allowed examination of emotion specificity), and instructions asked for best rather than actual responses (since it was assumed that EI was an ability rather than behavioural tendency). Alternatives to each of these three decisions would certainly have been possible, and the following paragraphs discuss the likely consequences of such alternatives and their utility in further research in this area.

Limiting the length of test items meant that subtleties and nuances of meaning in a situation may have been lost. Real emotional situations may be fragmented and confused with multiple sources of irrelevant information, such that ecologically valid items would mirror this complexity as closely as possible. The Emotional Accuracy Research Scale (EARS; Mayer & Geher, 1996; Geher et al., 2001) provides a close analogue to this ecologically valid alternative, with several relatively detailed descriptions of a situation constituting each item. However, this scale had very low internal consistency, particularly for a long (96-item) test, indicating that such complexity may result in a large amount of measurement error.

The relatively narrow number of emotions covered (fear, anger, and sadness only) and the exclusion of items containing more than one emotion also limited what the STEM assessed. Given Chapter 6's indication that anger-, fear- and sadness-management formed relatively distinct factors; it may be that management of other emotions (e.g., guilt-management) also differ considerably from sadness, anger, or fear management. Group differences, personality correlates, and predictive power may well vary for the management of different emotions.

Whether test items ask for actual behaviour rather than the ideal response may drastically change the construct measured. The *best response* (what action would be most effective for protagonist X?) instructions for the STEM were chosen on the assumption that emotion management skill was fundamentally an ability rather than a behavioural disposition. Nguyen and McDaniel (2003) found that best response instructions resulted in higher correlations with cognitive ability than actual response instructions, and Study 2 of Chapter 2 of this thesis found that free responses of actual behaviours were more strongly related to personality than those free responses of ideal or best behaviours. Contrasting the STEM with Freudenthaler and Neubauer's (2005) emotion management tests illustrates the likely consequences of asking for actual rather than best responses. Freudenthaler and Neubauer's situations and response options were very similar to the STEM, but test-takers were instructed to select the alternative that would best describe their own behaviour. Under these conditions measures correlated as strongly with personality as they did with each other (up to $r = .67$ with Neuroticism) and were virtually unrelated to cognitive ability. In contrast, the STEM was unrelated to Neuroticism (although weakly related to Agreeableness and Openness) and related to both Gf and Gc. Large construct-related differences may result from what appears to be small methodological differences.

One further possible alternative in the development of situational tests is the use of multi-media technology (video, audio or animated footage) rather than a textual presentation of item content. Chan and Schmitt (1997) found several advantages of using video-based measures over paper and pencil measures of work habits and interpersonal skills. These advantages included less confounding of the constructs of interest with reading comprehension, less adverse impact (in terms of a Black-White difference), and more favourable face-validity perceptions by test-takers (which may affect performance through increasing motivation of the test-takers). When investigating the overlap between EI and intelligence, the lack of reading-comprehension-related variance in scores may be particularly important: when both items and response options are *not* presented in text-based

format, then any overlap with text-based intelligence tests is more likely to be construct-related rather than method-related.

6. Concluding Comments

New theoretical and procedural paradigms such as appraisal theory and the SJT methodology were demonstrated to be useful approaches to the construction of tests assessing socio-emotional competencies. Despite different theoretical underpinnings, the Understanding component of Strategic EI clearly constitutes part of crystallized intelligence, with small distinctions from other Gc tests, such as the relationship to Agreeableness, and state anxiety. Management is relatively more distinct from existing conceptualisations of intelligence, although by no means independent from these constructs. The opposite pattern of gender differences for Management compared to many other intelligence test scores is intriguing and potentially useful. However, neither Management nor Understanding strongly predicted a wide range of positive life outcomes, relating only to a small number of criteria directly concerned with emotionality. The interest in, use of, and hype surrounding emotional intelligence as a measurable entity seem out of proportion with the evidence of EI's distinctiveness, utility, and ultimate importance. The idea, cultural phenomenon, or zeitgeist of EI is currently stronger than the empirical reality. Finally, although EI is not the mountain of promise that its immense popularity might indicate, the molehill of empirically-supported conclusions may have its uses: Slightly different group differences as well as differences in personality and emotionality-related correlates might potentially be used to lower adverse impact, and predict specific emotion-related outcomes.

Reference List

Abad, F. J., Colom, R., Juan-Espinosa, M., & Garcia, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence, 31*, 157-166.normal

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93*, 797-825.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychology Bulletin, 121*, 219-245.

Allspach, J. R., & Breining, K. (2005) Gender differences and trends over time for the SAT reasoning test. Retrieved 23 November 2005, from http://ftp.ets.org/pub/res/achgapallspach.pdf

Amthauer, R., Brocke, B., Leipmann, D., & Beauducel, A. (2000). Intelligenz-struktur-test. Gottingen: Hogrefe.

Anastasi, A. (1970). On the formation of psychological traits. *American Psychologist, 25*, 899-910.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*, 176-181.

Ashton, M. C., Lee, K., Vernon, P. A., & Jang, K. L. (2000). Fluid intelligence, crystallized intelligence, and the openness/intellect factor. *Journal of Research in Personality, 34*, 198-207.

Austin, E. J. (2004). An investigation of the relationship between trait emotional intelligence and emotional task performance. *Personality & Individual Differences, 36*, 1855-1864.

Austin, E. J., Saklofske, D. H., & Egan, V. (2005). Personality, well-being and health correlates of trait emotional intelligence. *Personality & Individual Differences, 38*, 547-558.

Austin, E. J., Saklofske, D. H., Huang, S. H. S., & McKenney, D. (2004). Measurement of trait emotional intelligence: Testing and cross-validating a modified version of Schutte et al.'s (1998) measure. *Personality & Individual Differences, 36*, 555-562.

Australian Bureau of Statistics. *Australia now*. Retrieved 11 July, 2005, from http://www.abs.gov.au/ausstats/abs@.nsf/2.6.1.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994a). The twenty-item Toronto Alexithymia Scale: I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*, 23-32.

Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994b). The twenty-item Toronto Alexithymia Scale: II. Convergent, discriminant and concurrent validity. *Journal of Psychosomatic Research, 38*, 33-40.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality & Social Psychology, 70*, 614-636.

Bar-On, R. (2000). Emotional and social intelligence: Insights from the emotional quotient inventory. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence*. San Francisco: Jossey-Bass.

Barchard, K. A. (2003). Does emotional intelligence assist in the prediction of academic success? *Educational & Psychological Measurement, 63*, 840-858.

Barchard, K. A., & Hakstian, R. A. (2004). The nature and measurement of emotional intelligence abilities: Basic dimensions and their relationships with other cognitive ability and personality variables. *Educational & Psychological Measurement, 64*, 437-462.

Bastian, V., Burns, N., & Nettelbeck, T. (2005). Emotional intelligence predicts life skills, but not as well as personality and cognitive abilities. *Personality & Individual Differences, 39*, 1135-1145.

Bates, T. C., & Shieles, A. (2003). Crystallized intelligence as product of speed and drive for experience: The relationship of inspection time and openness to g and Gc. *Intelligence, 31*, 275-287.

Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin Model of Intelligence Structure (BIS). European Journal of Psychological Assessment, 18, 97-112.

Birney, D. P., & Bowman, D. B. (in preparation). Dissociating working-memory and fluid-reasoning demand in the Graeco-Latin square task.

Boyatzis, R. E., Goleman, D., & Rhee, K. S. (2000). Clustering competence in emotional intelligence: Insights from the emotional competence inventory. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 343-362). San Fransisco: Jossey-Bass.

Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality & Social Psychology Bulletin, 29*, 1147-1158.

Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality & Individual Differences, 36*, 1387-1402.

Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry, 15*, 234-238.

Burt, C. (1954). The differentiation of intellectual ability. *British Journal of Educational Psychology, 24*, 76–90.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Caruso, D. R., Mayer, J. D., & Salovey, P. (2002). Relation of an ability measure of emotional intelligence to personality. *Journal of Personality Assessment, 79*, 306-320.

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality & Social Psychology*, 267-283.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.

Chan, D. W. (1995). The two scaled versions of the Chinese General Health Questionnaire: A comparative analysis. *Social Psychiatry & Psychiatric Epidemiology, 30*, 85-91.

Ciarrochi, J., Chan, A. Y. C., & Bajgar, J. (2001). Measuring emotional intelligence in adolescents. *Personality & Individual Differences, 31*, 1105-1119.

Ciarrochi, J., Deane, F. P., & Anderson, S. (2002). Emotional intelligence moderates the relationship between stress and mental health. *Personality & Individual Differences, 32*, 197-209.

Ciarrochi, J., Scott, G., Deane, F. P., & Heaven, P. C. L. (2003). Relations between social and emotional competence and mental health: A construct validation study. *Personality & Individual Differences, 35*, 1947–1963.

Ciarrochi, J. V., Chan, A. Y. C., & Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality & Individual Differences, 28*, 539-561.

Coffey, E., Berenbaum, H., & Kerns, J. G. (2003). The dimensions of emotional intelligence, alexithymia, and mood awareness: Associations with personality and performance on the emotional stroop. *Cognition & Emotion, 17*, 671-679.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Colom, R., Escorial, S., & Rebollo, I. (2004). Sex differences on the progressive matrices are influenced by sex differences on spatial ability. *Personality & Individual Differences, 37*, 1289-1293.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality & Social Psychology, 81*, 322-331.

Costa, P. T., Jr., & McCrae, R. R. (1980). Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. J. Brim (Eds.), *Life span development and behavior* (pp. 65-102). New York: Academic Press.

Costa, P. T., Jr., & McCrae, R. R. (1991). Four ways five factors are basic. *Personality & Individual Differences, 13*, 653-665.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFM) Professional Manual*. Odessa: FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64*, 21-50.

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal & Social Psychology, 60*, 151-174.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*, 111-131.

Cureton, E. E. (1950). Validity, reliability and baloney. *Educational & Psychological Measurement, 10*, 94-96.

Danthiir, V., Roberts, R. D., Pallier, G., & Stankov, L. (2001). What the nose knows: Olfaction and cognitive abilities. *Intelligence, 29*, 337-361.

Danthiir, V., Wilhelm, O., & Schacht. (submitted). Correct decision speed: Correctly an ability?

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

Davis, P. J. (1999). Gender differences in autobiographical memory for childhood emotional experiences. *Journal of Personality & Social Psychology, 76*, 498-510.

Dawda, D., & Hart, S. D. (2000). Assessing emotional intelligence: Reliability and validity of the Bar-On Emotional Quotient Inventory (EQ-i) in university students. *Personality & Individual Differences, 28*, 797-812.

Day, A. L., & Carroll, S. A. (2004). Using an ability-based measure of emotional intelligence to predict individual performance, group performance, and group citizenship behaviours. *Personality & Individual Differences, 36*, 1443-1458.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*, 71-75.

Duffy, C. J., Cunningham, E. G., & Moore, S. M. (2005) Brief report: The factor structure of mood states in an early adolescent sample. *Journal of Adolescence, 28*, 677-680

Ekman, P. (1992a). Are there basic emotions? *Psychological Review, 99*, 550-553.

Ekman, P. (1992b). An argument for basic emotions. *Cognition & Emotion, 6*, 169-200.

Ekman, P. (2004). *Microexpression Training Tools (METT) and Subtle Expression Training Tools (SETT)*. Information available at http://www.paulekman.com/research_cds.php.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*. Palo Alto: CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Oxford: Pergamon Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K. et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality & Social Psychology, 53*, 712-717.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion: An old controversy and new findings. In U. C. Segerstrale & P. Molnar (Eds.), *Nonverbal communication: Where nature meets culture* (pp. 27-46). Mahwah, NJ: Lawrence Erlbaum Associates.

Ekman, P., & Rosenberg, E., L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. New York: Oxford University Press.

Engelberg, E., & Sjoberg, L. (2004). Emotional intelligence, affect intensity, and social adjustment. *Personality & Individual Differences, 37*, 533-542.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*, 429-456.

Fitness, J. (2000). Anger in the workplace: An emotion script approach to anger episodes between workers and their superiors, co-workers and subordinates. *Journal of Organizational Behavior, 21*, 147-162.

Fitness, J., & Fletcher, G. J. O. (1993). Love, hate, anger and jealousy in close relationships: A prototype and cognitive appraisal analysis. *Journal of Personality & Social Psychology, 5*, 942-958.

Folkman, S., & Lazarus, R. S. (1980). An analysis of coping in a middle-aged community sample. *Journal of Health & Social Behavior, 21*, 219-239.

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality & Social Psychology, 48*, 150-170.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). Kit of reference tests for cognitive factors. Princeton, NJ: Educational Testing Service.

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

Frijda, N. H., & Zeelenberg, M. (2001). Appraisal: What is the dependent? In K. R. Scherer & A. Schorr (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 141-155). New York: Oxford University Press.

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case for emotional intensity. *Journal of Personality & Social Psychology, 61*, 427-434.

Funke, U., & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection & Assessment, 6*, 115-123.

Furnham, A. (2005, July). Opening address, *12th Biennial Meeting of the International Society for the Study of Individual Differences*. Adelaide, Australia.

Furnham, A., & Petrides, K. V. (2003). Trait emotional intelligence and happiness. *Social Behavior & Personality, 31*, 815-824.

Geher, G., Warner, R. M., & Brown, A. S. (2001). Predictive validity of the Emotional Accuracy Research Scale. *Intelligence, 29*, 373-388.

Gignac, G., Palmer, B., Manocha, R., & Stough, C. (2005). An examination of the factor structure of the Schutte Self-Report Emotional Intelligence (SSREI) scale via confirmatory factor analysis. *Personality & Individual Differences, 39*, 1029-1042

Gignac, G. E. (2005). Openness to experience, general intelligence and crystallized intelligence: A methodological extension. *Intelligence, 33*, 161-167.

Gignac, G. E., Stough, C., & Loukomitis, S. (2004). Openness, intelligence, and self-report intelligence. *Intelligence, 32*, 133-143.

Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire; a technique for the identification and assessment of non-psychotic psychiatric illness*. New York: Oxford University Press.

Goldberg, D. P. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality & Social Psychology, 59*, 1216-1229.

Goldberg, L. R. (2004). International personality item pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences. Retrieved 19 July, 2004, 2004, from http://www.ipip.ori.org.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam Books.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. J. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*, 504-528.

Gregory, R. J. (2004). *Psychological testing: History, principles, and applications*. Needham Heights, MA: Allyn & Bacon.

Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence, 29*, 523-535.

Guilford, J. P., Hendricks, M., & Hoepner, R. (1968). Solving social problems creatively. *Journal of Creative Behavior, 2*, 155-164.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*, 229-246.

Hanson, M. A., Horgen, K. E., & Borman, W. C. (1998). *Situational judgment: An alternative approach to selection test development.* Paper presented at the International Military Testing Association Conference, Pensacola, Florida.

Harris, J. A. (2004). Measured intelligence, achievement, openness to experience, and creativity. *Personality & Individual Differences, 36*, 913-929.

Hays, R. D., & DiMatteo, M. R. (1987). A short-form measure of loneliness. *Journal of Personality Assessment, 51*, 69-81.

Hedge, J. W., Bruskiewicz, K. T., Borman, W. C., Hanson, M. A., & Logan, K. K. (2000). Selecting pilots with crew resource management skills. *International Journal of Aviation Psychology, 10*, 377-392.

Hedlund, J., Forsythe, G. B., Horvath, J. A., Williams, W. M., Snook, S., & Sternberg, R. J. (2003). Identifying and assessing tacit knowledge: Understanding the practical intelligence of military leaders. *Leadership Quarterly, 14*, 117-140.

Hedlund, J., & Sternberg, R. J. (2000). Too many intelligences? Integrating social, emotional, and practical intelligence. In *Handbook of emotional intelligence* (pp. 136-167). San Fransisco: Jossey-Bass.

Henry, J. D., Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*, 227-239.

Herlitz, A., Nilsson, L.-G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*, 801-811.

Hexel, M. (2003). Alexithymia and attachment style in relation to locus of control. *Personality & Individual Differences, 35*, 1261-1270.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*, 253- 270.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford Press.

Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence, 6*, 165-185.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99*, 561-565.

Izard, C. E. (1993). Organizational and motivational functions of discrete emotions. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 631-641). NY: Guilford Press.

Izard, C. E. (2001). Emotional intelligence or adaptive emotions? *Emotion, 1*, 249-257.

Jacobsen, B. K., Hasvold, T., Hoyer, G., & Hansen, V. (1995). The general health questionnaire: How many items are really necessary in population surveys? *Psychological Medicine, 25*, 957-962.

Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CN: Praeger Publishing.

Johnson, J., A. (1991). Screening massively large data sets for non-responsiveness in web-based personality inventories. Retrieved 19 July, 2004, from http://www.personal.psu.edu/~j5j/papers/screening.html.

Johnson, J., A. (2004). Short form for the IPIP-NEO. Retrieved 19th July 2004, from http://www.personal.psu.edu/faculty/j/5/j5j/IPIP/ipipneo120.htm.

Kafetsios, K. (2004). Attachment and emotional intelligence abilities across the life course. *Personality & Individual Differences, 37*, 129-145.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier and Boyle (2005). *Psychological Bulletin, 131*, 66-71.

Kaufman, A. S., & Kaufman, J. C. (2001). Emotional intelligence as an aspect of general intelligence: What would David Wechsler say? *Emotion, 1*, 258-264.

Khavarpour, F., & Rissel, C. (1997). Mental health status of Iranian migrants in Sydney. *Australian & New Zealand Journal of Psychiatry, 31*, 823-834.

Kihlstrom, J. F., & Cantor, N. (2000). Social intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 359-379). New York: Cambridge University Press.

Kleitman, S., & Stankov, L. (2005). Self-confidence and meta-cognitive processes. *Journal of the Institute for Educational Research, 37*, 45-69.

Kline, T., J. B. (1994). Measurement of tactical and strategic decision making. *Educational & Psychological Measurement, 54*, 745-756.

Kline, T., J. B. (1996). A task for use in decision-making research and practice. *Personality & Individual Differences, 21*, 1077-1079.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*, 389-433.

Laird, R. (2005). What is it we think we are trying to fix and how should we fix it? A view from the admissions office. In W. J. Camara & E. W. Kimmel (Eds.), Choosing students: Higher education admissions tools for the 21st century (pp. 13-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Lam, L. T., & Kirby, S. L. (2002). Is emotional intelligence an advantage? An exploration of the impact of emotional and general intelligence on individual performance. *Journal of Social Psychology, 142*(1), 133-143.

Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *Journal of Organizational Behavior, 26*, 411-424.

Landy, F. J. (in press). The long, frustrating and fruitless search for social intelligence: A cautionary tale. In K. R. Murphy (Ed.), *The emotional intelligence bandwagon: The struggle between science and marketing for the soul of EI*. Mahwah: NJ: Lawrence Erlbaum Associates.

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.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lazarus, R. R. (1999). *Stress and emotion: A new synthesis*. New York: Springer.

Legree, P. J. (1995). Evidence for an oblique social intelligence factor established with a likert-based testing procedure. *Intelligence, 21*, 247-266.

Legree, P. J., Martin, D. E., & Psotka, J. (2000). Measuring cognitive aptitude using unobtrusive knowledge tests: A new survey technology. *Intelligence, 28*, 291-308.

Legree, P. J., Psotka, J., Tremble, T., & Bourne, D. R. (2004). Using consensus based measurement to assess emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *International handbook of emotional intelligence* (pp. 155-180). Göttingen, Germany: Hogrefe & Huber Publishers.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442-452.

Lopes, P. N., Brackett, M. A., Nezlek, J., B., Schutz, A., Sellin, I., & Salovey, P. (2004). Emotional intelligence and social interaction. *Personality and Social Psychology Bulletin, 30*, 1018-1034.

Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality & Individual Differences, 35*, 641-658.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research & Therapy, 33*, 335-343.

Lynn, R., Irwing, P., & Cammock, T. (2001). Sex differences in general knowledge. *Intelligence, 30*, 27-39.

Lyons, J. B., & Schneider, T. R. (2005). The influence of emotional intelligence on performance. *Personality & Individual Differences, 39*, 693-703.

MacCann, C., Roberts, R. D., Matthews, G., & Zeidner, M. (2003). Psychological assessment of emotional intelligence: A review of self-report and performance-based testing. *International Journal of Organizational Analysis, 11*, 247-274.

MacCann, C., Roberts, R. D., Matthews, G., & Zeidner, M. (2004). Consensus scoring and empirical option weighting of performance-based emotional intelligence (EI) tests. *Personality & Individual Differences, 36*, 645-662.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P. et al. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior, 24*, 179-209.

Matthews, G., Emo, A., Funke, G., Zeidner, M., Roberts, R. D., & Schulze, R. (submitted). Emotional intelligence, personality, and task-induced stress. *Journal of Experimental Psychology: Applied*.

Matthews, G., & Zeidner, M. (2000). Emotional intelligence, adaptation to stressful encounters, and health outcomes. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 459-789). San Fransisco: Jossey-Bass.

Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and myth*. Cambridge: MIT Press.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2000a). Selecting a measure of emotional intelligence: The case for ability scales. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 320-342). San Fransisco: Jossey-Bass.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2000b). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298.

Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22*, 89-114.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and EI: Educational implications* (pp. 3-34). New York: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. R. (1997). Multifactor emotional intelligence scale.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002a). *Mayer-Salovey-Caruso emotional intelligence test (*MSCEIT*) user's manual*. Toronto, Canada: MHS Publishers.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002b). *Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) item booklet*. Toronto, Canada: MHS Publishers.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 1*, 232-242.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT v2.0. *Emotion, 3*, 97-105.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection & Assessment, 9*, 103-113.

McDaniel, M. A., & Whetzel, D., L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence, 33*, 515-525

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: NJ: Lawrence Erlbaum Associates.

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64-82.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-179). New York: Guilford Press.

McGrew, K. S. (12 June, 2006). Cattell-Horn-Carroll CHC (Gf-Gc) theory: Past, present & future, version 74. from http://www.iapsych.com/CHCPP/CHCPP.HTML.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

McMillan, J., & Jones, F. L. (2000). The ANU3_2 scale: A revised occupational status scale for Australia. *Journal of Sociology, 36*, 64-80.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality & Social Psychology Bulletin, 27*, 769-783.

Montebarocci, O., Codispoti, M., Baldaro, B., & Rossi, N. (2004). Adult attachment style and alexithymia. *Personality & Individual Differences, 36*, 499-507.

Moss, F. A., Hunt, T., Omwake, K. T., & Woodward, L. G. (1955). *Manual for the George Washington University series social intelligence test*. Washington, D.C: Center for Psychological Service.

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.

Nemiah, J. C., Freyberger, H., & Sifneos, P. E. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (Ed.), *Modern trends in psychosomatic medicine* (Vol. 3, pp. 430-439). London: Butterworths.

Newsome, S., Day, A. L., & Catano, V. M. (2000). Assessing the predictive validity of emotional intelligence. *Personality & Individual Differences, 29*, 1005-1016.

Nguyen, N. T., & McDaniel, M. A. (2003). Response instructions and racial differences in a situational judgment test. *Applied H.R.M. Research, 8*, 33-44.

Roberts, R. D., Schultze, R., O'Brien, K., Reid, J., MacCann, C., & Maul, A., (submitted). Where are the emotions in emotional intelligence measures? *Emotion*.

O'Conner, R. M., & Little, I. (2003). Revisiting the predictive validity of emotional intelligence: Self-report versus ability-based measures. *Personality & Individual Differences, 34*, 1-10.

O'Sullivan, M., & Ekman, P. (2004). Facial expression recognition and emotional intelligence. In G. Geher (Ed.), *Measuring emotional intelligence* (pp. 89-109). NY: Nova Science Publishing.

O'Sullivan, M., & Guilford, J. P. (1975). Six factors of behavioral cognition: Understanding other people. *Journal of Educational Measurement*, *12*, 255-271.

O'Sullivan, M., & Guilford, J. P. (1976). *Four factor tests of social intelligence (behavioral cognition): Manual of instructions and interpretations*. Orange, CA: Sheridan Psychological Services.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition & Emotion, 1*, 29-50.

Oberauer, K., Schulze, R., Wilhelm, O., & Suess, H.-M. (2005). Working memory and intelligence - their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61-65.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*, 315-331.

Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles, 5*, 265-276.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L. et al. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology, 129*, 257-299.

Palmer, B., Donaldson, C., & Stough, C. (2002). Emotional intelligence and life satisfaction. *Personality & Individual Differences, 33*, 1091-1100.

Palmer, B., Gignac, G., Manocha, R., & Stough, C. (2005). A psychometric evaluation of the Mayer-Salovey-Caruso Emotional Intelligence Test version 2.0. *Intelligence, 33*, 285-305.

Panksepp, J. (1992). A critical role for "affective neuroscience" in resolving what is basic about basic emotions. *Psychological Review, 99*, 554-560.

Parker, J. D. A. (2004). The relevance of emotional intelligence for clinical psychology. In R. Schulze & R. D. Roberts (Eds.), *International handbook of emotional intelligence* (pp. 271-287). Göttingen, Germany: Hogrefe & Huber Publisher.

Parker, P. D., Prkachin, K. M., & Prkachin, G. C. (2005). Processing of facial expressions of negative emotion in alexithymia: The influence of temporal constraint. Journal of Personality, 73, 1087-1107.

Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2004). Emotional intelligence and academic success: Examining the transition from high school to university. *Personality & Individual Differences*, 163-172.

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment, 5*, 164-172.

Perez, J. C., Petrides, K. V., & Furnham, A. (2005). Measuring trait emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *International handbook of emotional intelligence* (pp. 181-202). Göttingen: Germany: Hogrefe & Huber Publishers.

Petrides, K. V., Frederickson, N., & Furnham, A. (2004). The role of trait emotional intelligence in academic performance and deviant behavior at school. *Personality & Individual Differences, 36*, 277-293.

Petrides, K. V., & Furnham, A. (2000). On the dimensional structure of emotional intelligence. *Personality & Individual Differences, 29*, 313-320.

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15*, 425-448.

Petrides, K. V., & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality, 17*, 39–57.

Ployhart, R., E., & Ehrhart, M., G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tasks. *International Journal of Selection & Assessment, 11*, 1-16.

Plutchik, R. (1997). The circumplex as a general model of the structure of emotions and personality. In R. Plutchik & H. R. Conte (Eds.), *Circumplex models of personality and emotions* (pp. 17-45). Washington, DC: American Psychological Association.

Raaijmakers, Q. A. W., van Hoof, A., Hart, H., Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on likert-type scale items: Neutral or missing values? *International Journal of Public Opinion Research, 12*, 208-216.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Sec 4. Advanced progressive matrices*. Oxford: Oxford Psychologists Press.

Reid, J. (2005, July). Using a criterion to provide a foundation for EI test development. Paper presented at the *12th Biennial Meeting of the International Society for the Study of Individual Differences*, Adelaide, Australia.

Riggio, R. E., Messamer, J., & Throckmorton, B. (1991). Social and academic intelligence: Conceptually distinct but overlapping constructs. *Personality & Individual Differences, 12*(695-702).

Roberts, R. D. (2000). *Openness Conscientiousness Extraversion Agreeableness Neuroticism index condensed (OCEANIC) test manual*. Sydney, Australia: E-ntelligence Testing Products.

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB): Little more than acculturated learning (Gc)! *Learning & Individual Differences, 12*, 81-103.

Roberts, R. D., Schultze, R., Zeidner, M., & Matthews, G. (2004). Understanding, measuring and applying emotional intelligence: What have we learned? What have we missed? In R. Schultze & R. D. Roberts (Eds.), *International handbook of emotional intelligence* (pp. 311-341). Göttingen, Germany: Hogrefe & Huber.

Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning & Individual Differences, 11*, 1-120.

Roberts, R. D., Stankov, L., Pallier, G., & Dolph, B. (1997). Charting the cognitive sphere: Tactile-kinesthetic performance within the structure of intelligence. *Intelligence, 25*, 111-148.

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion, 1*, 196-231.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of Personality and Social Psychology, 5*, 11-36.

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition & Emotion, 5*, 161-200.

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. Series in affective science* (pp. 68-91). New York: Oxford University Press.

Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion, 10*, 241-277.

Roseman, I. J., Dhawan, N., Rettek, S., Naidu, R. K., & Thapa, K. (1995). Cultural differences and cross-cultural similarities in appraisals and emotional responses. *Journal of Cross-Cultural Psychology, 26*, 23-48.

Roseman, I. J., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition & Emotion, 18*, 1-28.

Roseman, I. J., Spindel, M. S., & Jose, P. E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality & Social Psychology, 5*, 899-915.

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality & Social Psychology, 67*, 206-221.

Ross, M., & Holmberg, D. (1992). Are wives' memories for events in relationships more vivid than their husbands' memories? *Journal of Social & Personal Relationships, 9*, 585-604.

Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal, 48*, 845-858.

Russell, D., Peplau, L. A., & Cutrona, L. E. (1980). The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. *Journal of Personality & Social Psychology, 39*, 472-480.

Saarni, C. (1999). *The development of emotional competence*. New York: Guilford Press.

Saklofske, D. H., Austin, E. J., & Minski, P. S. (2003). Factor structure and validity of a trait emotional intelligence measure. *Personality & Individual Differences, 34*, 707–721.

Salovey, P., Bedell, B. T., Detweiller, J. B., & Mayer, J. D. (2000). Coping intelligently: Emotional intelligence and the coping process. In C. R. Snyder (Ed.), *Coping: The psychology of what works* (pp. 141-164). New York: Oxford University Press.

Salovey, P., & Mayer, J. D. (1989). Emotional intelligence. *Imagination, Cognition & Personality, 9*, 185-211.

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the trait meta-mood scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, & health* (pp. 125-154). Washington, DC: American Psychological Association.

Salovey, P., Stroud, L. R., Woolery, A., & Epel, E. S. (2002). Perceived emotional intelligence, stress reactivity, and symptom reports: Further explorations using the trait meta-mood scale. *Psychology & Health, 17*, 611-627.

Schaie, K.W. (1994). The course of adult intellectual development. *American Psychologist, 49*, 304-313

Scherer, K. R. (1993). Studying the emotion-antecedent appraisal processes: An expert systems approach. *Cognition & Emotion, 7*, 325-355.

Scherer, K. R. (1999a). Appraisal theory. In T. Dalgleish, T. & M. J. Power (Eds.). *Handbook of cognition and emotion.* (pp. 637-663). Chichester, England: John Wiley & Sons.

Scherer, K. R. (1999b). On the sequential nature of appraisal processes: Indirect evidence from a recognition task. *Cognition & Emotion, 13*, 763-793.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (pp. 92-120). New York: Oxford University Press.

Scherer, K. R. (Ed.). (1988). *Facets of emotion: Recent research*: Hillsdale, NJ: Lawrence Erlbaum Associates.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76-92.

Scherer, K. R., & Tannenbaum, P. H. (1986). Emotional experiences in everyday life: A survey approach. *Motivation & Emotion*, 295-314.

Schorr, A. (2001). Appraisal: The evolution of an idea. In K. R. Scherer & A. Schorr (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 20-34). New York: Oxford University Press.

Schulte, M. J., Ree, M. J., & Carretta, T. R. (2004). Emotional intelligence: Not much more than g and personality. *Personality & Individual Differences, 37*, 1059-1068.

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

Seidlitz, L., & Diener, E. (1998). Sex differences in the recall of affective experiences. *Journal of Personality & Social Psychology, 74*, 262-271.

Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Longterm retest and observer ratings. *Personnel Psychology, 39*, 791–809.

Sifneos, P. E. (1973). The prevalence of “alexithymic” characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22*, 255-262.

Slaski, M., & Cartwright, S. (2002). Health, performance and emotional intelligence: An exploratory study of retail managers. *Stress & Health, 18*, 63-68.

Slaski, M., & Cartwright, S. (2003). Emotional intelligence training and its implications for stress, health and performance. *Stress & Health, 19*, 233-239.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality & Social Psychology, 48*, 813-838.

Snekkevik, H., Anke, A. G. W., Stanghelle, J. K., & Fugl-Meyer, A. R. (2004). Is sense of coherence stable after multiple trauma? *Clinical Rehabilitation, 17*, 443-453.

Spearman, C. (1927). *The abilities of man*. New York: MacMillan.

Stankov, L. (1997). *Gf–Gc quickie test battery*. Sydney, Australia: E-ntelligence Testing Products.

Stankov, L. (2000a). Complexity, metacognition, and fluid intelligence. *Intelligence, 28*, 121-143.

Stankov, L. (2000b). Structural extensions of a hierarchical view on human cognitive abilities. *Learning & Individual Differences, 12*, 35-51.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology, 72*, 21-44.

Stankov, L., Seizova-Cajic, T., & Roberts, R. D. (2001). Tactile and kinesthetic perceptual processes within the taxonomy of human cognitive abilities. *Intelligence, 29*, 1-29.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M. et al. (2000). *Practical intelligence in everyday life*: New York: Cambridge University Press.

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A. et al. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence, 29*, 401-418.

Taylor, G. J. (2001). Low emotional intelligence and mental illness. In J. Ciarrochi & J. P. Forgas (Eds.), *Emotional intelligence in everyday life: A scientific inquiry* (pp. 67-81). New York: Psychology Press.

Taylor, G. J., & Bagby, R. M. (2000). An overview of the alexithymia construct. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence (pp. 40-67).* San Francisco: Jossey-Bass.

Tomkins, S. S. (1968). Affects: Primary motives of man. *Humanitas, 3*, 321-345.

Tottenham, N. *NIMSTIM face stimulus set*. Retrieved 17th February, 2004, from http://www.macbrain.org/faces.

Trinidad, D. R., & Johnson, C. A. (2002). The association between emotional intelligence and early adolescent tobacco and alcohol use. *Personality & Individual Differences, 32*, 95-105.

Trinidad, D. R., Unger, J. B., Chou, C., & Johnson, C. A. (2004). The protective association of emotional intelligence with psychosocial smoking risk factors for adolescents. *Personality & Individual Differences, 36*, 945-954.

Turner, T. J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review, 99*, 566-571.

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

van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*, 346-360.

van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behaviour, 65*, 71-95.

Wagner, R. K., & Sternberg, R. J. (1987). Tacit knowledge in managerial success. *Journal of Business & Psychology, 1*, 301-312.

Wagner, R. K., Sujan, H., Sujan, M., Rashotte, C. A., & Sternberg, R. J. (1999). Tacit knowledge in sales. In R. J. Sternberg & J. A. Horvath (Eds.). *Tacit knowledge in professional practice: Researcher and practitioner perspectives.* (pp. 155-182). Mahwah, NJ: Lawrence Erlbaum Associates.

Warwick, J., & Nettelbeck, T. (2004). Emotional intelligence is.? *Personality & Individual Differences, 37*, 1091-1100.

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology. Volume 10: Assessment psychology (pp. 43-66)*. New York: John Wiley & Sons.

Weis, S., Seidel, K., & Suess, H.-M. (2005, July). Facets of social intelligence: New ways of assessment, paper presented at the *12th Biennial Meeting of the International Society for the Study of Individual Differences*, Adelaide, Australia.

Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational & Psychological Measurement, 64*, 956-972.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage Publications.

Wilhelm, O. (2004). Measures of emotional intelligence: Practice and standards. In R. Schulze & R. D. Roberts (Eds.), *International handbook of emotional intelligence* (pp. 131-154). Göttingen, Germany: Hogrefe & Huber Publishers.

Woitaszewski, S. A., & Aalsma, M. C. (2004). The contribution of emotional intelligence to the social and academic success of gifted adolescents as measured by the Multifactor Emotional Intelligence Scale - adolescent version. *Roeper Review, 27*, 25-30.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.

Zeidner, M., Matthews, G., & Roberts, R. D. (2001). Slow down, you move too fast: Emotional intelligence remains an "elusive" intelligence. *Emotion, 1*, 196–231.

Zeidner, M., Matthews, G., & Roberts, R. D. (in press). Emotional intelligence, adaptation, and coping. In J. F. J. Ciarrochi, & J. D. Mayer (Ed.), *Emotional intelligence in everyday life: A scientific inquiry* (2nd edition ed.). Philadelphia, PA: Psychology Press.

Zeidner, M., Matthews, G., Roberts, R. D., & MacCann, C. (2003). Development of emotional intelligence: Towards a multi-level investment model. *Human Development, 46*, 69-96.