

Emotional Intelligence as Ability:
Assessing the Construct Validity of Scores from the
Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)

Hallvard Føllesdal

Department of Psychology

Faculty of Social Sciences

University of Oslo

TABLE OF CONTENTS

<i>Acknowledgements</i>	<i>v</i>
<i>Preface</i>	<i>vii</i>
<i>Summary</i>	<i>ix</i>
<i>List of Papers</i>	<i>xi</i>
1 Background	1
1.1 What is EI?	2
1.1.1 EI as Mental Ability	2
1.1.2 EI as Mixed Models.....	5
1.1.2.1 The ECI	6
1.1.2.2 The BarOn EQ-i	6
1.2 The MSCEIT – a Performance Measure of EI	7
1.2.1 The Domain for the Four Branch Ability Model.....	7
1.2.2 The MSCEIT Measurement Design	8
1.2.3 The Consensus Method Determines Correct Answers	10
1.3 The Construct Validity of the Scores from MSCEIT	11
1.3.1 Assessing MSCEIT Scores from the Perspective of Generalizability Theory.....	12
1.3.1.1 Correspondence between Measurement Design and Estimation Procedure.....	13
1.3.1.2 Relevant Sources of Variance for Estimating Reliability	14
1.3.2 Assessing the External Aspect of Construct Validity of MSCEIT Scores.....	15
1.3.2.1 Empathy	16
1.3.2.2 Leader Effectiveness	17
1.3.2.3 Transformational Leadership.....	18
2 Aims of the Thesis	19
3 Methods	20

3.1	Sample and Procedure	20
3.2	Measurements	20
3.2.1	EI	21
3.2.2	Empathy.....	21
3.2.3	Leader Effectiveness	21
3.2.4	Transformational Leadership.....	22
3.2.5	Personality Traits.....	22
4	<i>Statistical Methods</i>	22
4.1	Generalizability Theory	23
4.2	Multilevel Confirmatory Factor Analysis	24
5	<i>Results</i>.....	26
5.1	Summary of Paper 1.....	26
5.2	Summary of Paper 2.....	27
5.3	Summary of Paper 3.....	28
6	<i>General Discussion</i>	30
6.1	The Construct Validity of Scores from MSCEIT	30
6.1.1	The Content Aspect of Construct Validity	30
6.1.1.1	The Items May not Represent all Important Aspects of the Domain	31
6.1.1.2	The Scoring Method May Be Inappropriate.....	32
6.1.2	The Substantive Aspect of Construct Validity	33
6.1.3	The Structural Aspect of Construct Validity	35
6.1.4	The Generalizability Aspect of Construct Validity	36
6.1.5	The External Aspect of Construct Validity.....	38
6.1.6	The Consequential Aspect of Construct Validity	39
6.2	The Scientific Contributions of These Findings.....	39

7	<i>Conclusion and Suggestions for Future Research</i>	40
8	<i>References</i>	42
9	<i>Papers I – III</i>	52

ACKNOWLEDGEMENTS

This project has been funded by the Norwegian Research Council and conducted at the Department of Psychology, University of Oslo. First of all, I want to thank my supervisor, Professor Knut A. Hagtvet, for sharing his expertise in research methodology, and for encouraging me to learn Generalizability Theory and Multi-level Confirmatory Factor Analysis. This has been a challenging pursuit, and I am very grateful for his support and expert advice throughout this process.

I would like to thank cand. psychol. Geir Kalstad for collaborating with me in translating the MSCEIT and for his help in making this project possible, the Administrative Research Foundation (AFF) at the Norwegian School of Economics and Business Administration (NHH), in particular Arne Kjøde, Per Einar Olsen, Atle Jordahl, and Henning Lampe-Olsen, for providing me access to many of the participants in this project, and Professor Geir Kaufmann for his valuable advice and assistance in the earlier stages of the project. I am also grateful to the leaders and their subordinates who completed the questionnaires for these studies and the psychology students who helped me collect data with the MSCEIT. Finally I want to thank cand. psychol. John Holmes for back-translating questionnaires to English.

This project has been exciting and intellectually challenging, but also very time consuming. I am grateful to my parents, siblings, friends and colleagues, for their support and encouragement.

Hallvard Føllesdal

Oslo, April, 2008

PREFACE

Since this thesis was submitted, a revised version of Paper 1 has been accepted for publication. The reference is:

Føllesdal, H., & Hagtvet, K. A., Emotional intelligence: The MSCEIT from the perspective of generalizability theory, *Intelligence* (2008), doi:10.1016/j.intell.2008.08.005.

SUMMARY

This thesis presents the results from three papers assessing the validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2002). The MSCEIT is the only performance test measuring the entire four-branch ability model of EI (Mayer & Salovey, 1997; Salovey & Mayer, 1990).

Previous studies have reported low reliability coefficients for the branch scores for MSCEIT and reliability estimates vary greatly from study to study. The reported reliability coefficients may be biased or inflated, however, as the many sources of variance in the MSCEIT measurement design have not been taken into account when the coefficients have been estimated. Mixed results have also been reported regarding the construct validity of scores. More studies are therefore needed on the validity of scores from MSCEIT.

In Paper 1, Generalizability Theory (GT) is used to estimate the relative magnitude of the many sources of variance in the measurement design and generalizability (reliability) coefficients for the scores, taking the many explicit sources of variance in scores into account. Participants were 111 leaders of various businesses in Norway. The results reveal some important sources of variance in the measurement design that has neither been specified in the MSCEIT measurement design nor in the theory of EI. Only some of the branches in the MSCEIT provide scores that are generalizable. Moreover, the task Faces (from the branch Perceiving Emotions) provide scores that constitute three correlated factors, representing interactions of type of emotion expressed by faces and type of emotion to be rated. Moreover, the three scales seem to assess ability to identify emotions that are *absent* in faces, rather than emotions that are *present* in faces. These results are not in accordance with the theory of EI, suggesting that the validity of the scores derived from MSCEIT need to be assessed further.

In Paper 2, the validity of scores from MSCEIT is further assessed by relating the scale scores derived in Paper 1 to subordinates ratings of empathy and leader effectiveness. The theory of EI suggest that EI should be positively related to both empathy and leader effectiveness. Previous studies using empathy as a validity criterion, however, have used *self-ratings* of empathy despite the validity of self-ratings of emotional abilities have been questioned. In Paper II, both self-ratings and subordinate ratings of empathy were used. Multilevel Confirmatory Factor Analysis (MCFA) with latent variables was used to analyze the scores from 104 leaders (same sample as in Paper 1) and 459 subordinates' ratings of empathy and leader effectiveness. Scores from MSCEIT were found to be unrelated to ratings of empathy and leader effectiveness, suggesting that the validity of the scores may be questioned.

Paper 3 assesses the validity of the scores from MSCEIT further, by relating them to subordinates' ratings of their leader's transformational leadership behavior. It has been suggested that EI should be positively related to transformational leadership. The results from MCFA of 459 subordinates' ratings of transformational leadership and scores from 104 leaders on MSCEIT (same scales and sample as in Paper 1 and Paper 2) revealed that scores from MSCEIT were unrelated to transformational leadership. Scores from Facilitation, however, was related to one transformational leadership factor, but this relationship was not significant after controlling for personality scores provided by the NEO PI-R.

Together, the results from the three papers question the validity of the scores from MSCEIT. These findings are important, as the MSCEIT is frequently used to measure EI in research and applied settings and is regarded as the test that provides the best available evidence for the four-branch theory of EI. The findings from this thesis thus question the validity of the most important source of evidence for the four-branch ability model of EI.

LIST OF PAPERS

The following papers constitute this thesis:

Paper I

Føllesdal, H., & Hagtvet, K. A. (2008). Emotional Intelligence as ability: The MSCEIT from the perspective of generalizability theory. *Manuscript submitted for publication.*

Paper II

Føllesdal, H., & Hagtvet, K. A. (2008). Assessing the validity of the Mayer, Salovey, and Caruso Emotional Intelligence Test: Predicting empathy and effectiveness among leaders. *Manuscript submitted for publication.*

Paper III

Føllesdal, H. & Hagtvet, K. A. (2008). Assessing the validity of the Mayer, Salovey, and Caruso Emotional Intelligence Test 2.0: Predicting transformational leadership. *Manuscript submitted for publication.*

1 BACKGROUND

There has been a great interest in the construct of emotional intelligence (EI) during recent years. During the last five years, 150-200 scholarly works related to EI have been referenced yearly in the bibliographic database PsycInfo (search term “emotional-intelligence”). A chapter is also devoted to EI in the Annual Review of Psychology for 2008 (Mayer, Roberts, & Barsade, 2008), a journal that aims to review significant developments in psychology. The scientific status and utility of the EI construct, however, is uncertain. This thesis will assess the validity of scores from a test that provides the best evidence available for the theory of EI.

Many different approaches to EI have been developed, but the Mayer-Salovey four-branch ability model (Mayer & Salovey, 1997; Salovey & Mayer, 1990) is often regarded as the most scientifically based approach, and many use it as a point of departure in research. The empirical evidence for this model has been provided by the Mayer, Salovey, & Caruso Emotional Intelligence Test (MSCEIT, Mayer et al., 2002), which presently is the only performance test assessing the four-branch model (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006, p. 791). The MSCEIT has been translated to over twenty languages and is frequently used in both research and applied settings.

The reliability and validity of scores provided by the MSCEIT, however, needs to be assessed more thoroughly for at least three reasons: First, several studies have reported low reliability coefficients for some of the branch scores. Second, the reported reliability coefficients in previous studies have been estimated with methods that do take into account the many sources of variance in the MSCEIT measurement design, which may lead to biased or even inflated estimates. Third, several studies have provided mixed results regarding the external validity of the scores, suggesting that more studies are needed on this issue.

The major aim of the present thesis is to assess the construct validity of the scores from the MSCEIT. The thesis consists of three papers. In Paper 1, the magnitude of the many sources of variance in the measurement design will be estimated along with generalizability (reliability) coefficients for the scores. In Paper 2 and Paper 3, the validity of the scores will be further assessed by relating them to ratings of empathy, effectiveness and transformational leadership behavior, variables that are assumed to be positively related to EI.

1.1 What is EI?

There exist many different conceptualizations or models of EI. Mayer et al. (2000) differentiate between two different types of models: the “ability model” and “mixed models”. In the ability model, EI is defined as a mental ability that meets traditional standards for intelligence (Mayer, Caruso, & Salovey, 2000). Mixed models, on the other hand, refer to models that also includes other characteristics than EI, e.g., personality traits (Mayer, Salovey et al., 2000). This thesis will assess the construct validity of the scores from the MSCEIT, which is the only performance test of the four-branch ability model. As the ability model is often set apart from mixed models, both the ability approach and the mixed models approach will be presented briefly in the following.

1.1.1 EI as Mental Ability

The first definition of EI to appear in a scientific journal was provided by Peter Salovey and Jack Mayer in 1990 (Salovey & Mayer, 1990). This definition was slightly revised in 1997 to the following:

Emotional intelligence involves the ability to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge; and the ability to regulate emotions to promote emotional and intellectual growth. (Mayer & Salovey, 1997, p. 10)

This approach is often referred to as the Mayer-Salovey four-branch ability model of EI. Mayer et al. (2000) argue that the construct of EI must meet three criteria in order to be considered an intelligence. First, it must meet a *conceptual* criterion, i.e., it must reflect *mental performance* rather than preferred ways of behaving, a person's self-esteem or non-intellectual attainments. Second, it must meet a *correlational* criterion, i.e., it should define a set of abilities that are moderately correlated with one another, with other types of intelligence, and with various external criteria. And third, it must meet a *developmental* criterion, which entails that EI must develop with age and experience. They provide evidence suggesting that scores from performance measures of EI satisfy these criteria (Mayer, Caruso et al., 2000). This thesis, however, will assess more thoroughly the reliability and validity of scores from the MSCEIT, which is the only test measuring the entire four-branch model of EI.

EI is assumed to be more validly assessed by measures based on performance rather than self-report (e.g., Mayer & Geher, 1996), as the latter may be distorted by various types of biases. For instance, for general intelligence, self-report measures have been found to correlate only moderately with scores provided by performance tests of intelligence, often in the range of .30 - .35 (Paulhus, Lysy, & Yik, 1998). When it comes to EI, however, scores from self-report seem to be weakly correlated, or even unrelated, to scores from performance tests of EI. For instance, self-report of ability to infer accurately what other people think and feel have been found to be unrelated to scores from performance tests of this ability

(Marangoni, Garcia, Ickes, & Teng, 1995). Therefore, “ability tests of emotional intelligence were developed in an effort to overcome the limitations of self-reports of emotional intelligence and provide a more objective assessment of people’s actual rather than self-perceived abilities” (Lopes, Côté, & Salovey, 2006, p. 58). Performance measures of EI are thus hypothesized to be better predictors of emotional intelligent behavior than self-report measures (Mayer & Geher, 1996).

While the MSCEIT is the only performance test that measures the entire four-factor ability model, several tests have been developed to measure ability to perceive emotions that are expressed by other people, one of the many abilities encompassed by the Perceiving Emotions branch in the four-factor model. Examples of such measures are the Diagnostic Analysis of Non-Verbal Accuracy (DANVA, Nowicki & Duke, 2001) and Matsumoto and Ekman’s Japanese and Caucasian Brief Affect Recognition Test (Matsumoto et al., 2000), which both are performance tests of this ability. In such tests, people are asked to identify the emotion expressed in pictures of faces, which may be presented for only a brief period of time. Other tests use videotaped naturalistic interpersonal interactions, e.g., as in measures of empathic accuracy (Ickes, Marangoni, & Garcia, 1997).

When it comes to the entire four-branch model of EI, the Multifactor Emotional Intelligence Scale (MEIS, Mayer, Caruso et al., 2000) was also designed to measure the four branch ability model, but factor analyses of the 12 task scores suggests that the scores are best represented by three factors, rather than four as proposed by the theory. Thus, the scores from MEIS do not support the entire four-factor ability model. The MSCEIT, however, measures the four branches by eight tasks, and factor analyses of the task scores allegedly support a four-factor structure corresponding to the theoretical model (Mayer, Salovey, Caruso, & Sitarenios, 2003). In this thesis, the reliability and validity of the scores from

MSCEIT will be assessed further. First, however, the mixed models approach will be presented, as the ability model is often set apart from this approach.

1.1.2 EI as Mixed Models

Many researchers use the term EI to refer to several characteristics not usually conceived of as intelligence, e.g., personality traits, despite the term ‘EI’ indicates a kind of intelligence. Salovey and Mayer therefore termed such models as “mixed models”, in order to differentiate them from ability-based models.

The mixed models approach emerged after the publication of Daniel Goleman’s book *Emotional Intelligence* (Goleman, 1995). The book became a best-seller, and has been cited frequently in the scholarly literature. In this and later books, Goleman claims that EI is very important for many aspects of life, e.g., that it is four times more important for success than general intelligence (Goleman, 1998, p. 53). Such claims may catch attention, as decades of research has shown the general intelligence is a good predictor of several outcomes, e.g., academic achievement and work performance (Schmidt & Hunter, 1998; Sternberg, Grigorenko, & Bundy, 2001). Many of Goleman’s claims, however, are not supported by evidence, and in addition, Goleman uses the term EI to refer to a wide array of different characteristics not usually conceived of as ability, such as personality traits. Some argue that Goleman uses the term EI to refer to all positive characteristics that are not intelligence (Matthews, Zeidner, & Roberts, 2002, p. 12).

Several questionnaires have subsequently been developed to measure the many characteristics various researchers associate with the term EI. Different models often encompass very different constructs, and the models are usually referred to by the name of the questionnaire developed to measure the model. The two most well-known models are the

Emotional Competence Inventory (ECI, Boyatzis, Goleman, & Rhee, 2000) and the BarOn Emotional Quotient Inventory (BarOn EQ-i, Bar-On, 1997).

1.1.2.1 The ECI

The ECI is a questionnaire developed by Daniel Goleman and Richard Boyatzis (Boyatzis et al., 2000) that measures 18 competencies grouped into four broad categories: self-awareness, self-management, social awareness, and social skills (Sala, 2006). Among the competencies measured are characteristics such as emotional self-awareness, empathy, conscientiousness and conflict management. Boyatzis and Goleman (2000) define EI as follows: “Emotional intelligence is observed when a person demonstrates the competencies that constitute self-awareness, self-management, social awareness, and social skills at appropriate times and ways in sufficient frequency to be effective in the situation” (p. 344). The ECI measures these competencies by self-ratings or by other ratings of the extent to which various behaviors are characteristic of the person in question. Some argue that the questionnaire measures mostly personality traits and only to a small extent EI (Matthews et al., 2002; Mayer, Salovey et al., 2000). It has also been questioned to what extent the questionnaire measures anything else than traditional measures of leadership behavior (Byrne, Dominick, Smither, & Reilly, 2007).

1.1.2.2 The BarOn EQ-i

Another well-known mixed model of EI is the BarOn EQ-i (Bar-On, 1997). Here, EI is defined as “an array of noncognitive capabilities, competencies, and skills that influence one’s ability to succeed in coping with environmental demands and pressures” (Bar-On, 1997, p. 14). This definition of EI is thus very different from Mayer and Salovey’s definition, where EI refers to *mental* or *cognitive* abilities. The BarOn EQ-i is a self-report questionnaire

measuring 15 characteristics categorized into five broad areas. The scores, however, have been found to correlate highly with scores from personality measures like the NEO PI-R (Bar-On, 2000), and it has been argued that the EQ-i to a large extent measures personality traits, not EI (Mayer, Salovey et al., 2000). A questionnaire has also been developed to measure these characteristics by other peoples ratings, the BarOn EQ360 (Bar-On, 2002). In Paper 2, the empathy scales from both the BarOn EQ-i and the BarOn EQ360 will be used, as empathy is a characteristic assumed to be strongly related to EI.

1.2 The MSCEIT – a Performance Measure of EI

While several questionnaires have been developed to measure EI as mixed models, the MSCEIT is the only performance test measuring the entire four-branch *ability* model of EI. Thus, the MSCEIT provides the best evidence available for the four-branch ability model.

1.2.1 The Domain for the Four Branch Ability Model

The domain for the four-branch ability model consists of four sub-domains, or branches, each consisting of several abilities assumed to be related (Mayer & Salovey, 1997; Salovey & Mayer, 2002). The domain for the entire model is outlined in Table 1. Each branch is relatively broadly defined, and the abilities mentioned in Table 1 are only examples of the many abilities encompassed by each branch. For instance, in other descriptions of the first branch of the model (which is usually termed ‘Perceiving Emotions’) the branch encompasses “ability to perceive emotions in oneself and others, as well as in objects, art, stories, music, and other stimuli” (Mayer et al., 2002, p. 7). Thus, the domain is defined relatively broadly, as is also the case for the other branches in the model.

Table 1

The Domain of EI

Emotional Perception and Expression

- Ability to identify emotions in one's physical and psychological states
- Ability to identify emotion in other people
- Ability to express emotions accurately and to express needs related to them
- Ability to discriminate between accurate/honest and inaccurate/dishonest feelings

Emotional Facilitation of Thought (Using Emotional Intelligence)

- Ability to redirect and prioritize thinking on the basis of associated feelings
- Ability to generate emotions to facilitate judgment and memory
- Ability to capitalize on mood changes to appreciate multiple points of view
- Ability to use emotional states to facilitate problem-solving and creativity

Emotional Understanding

- Ability to understand relationships among various emotions
- Ability to perceive the causes and consequences of emotions
- Ability to understand complex feelings, emotional blends, and contradictory states
- Ability to understand transitions among emotions

Emotional Management

- Ability to be open to feelings, both pleasant and unpleasant
 - Ability to monitor and reflect on emotions
 - Ability to engage, prolong, or detach from an emotional state
 - Ability to manage emotions in oneself
 - Ability to manage emotions in others
-

Note. Adapted from Salovey and Mayer (2002).

1.2.2 The MSCEIT Measurement Design

The MSCEIT is designed to measure the entire four-branch model. The measurement design is complex and multifaceted (Table 2) and consists of four branches, corresponding to the four branches in the theoretical model: Perceiving Emotions, Using Emotions, Understanding Emotions, and Managing Emotions. Each branch is measured by two tasks, and each task measures the underlying abilities by a set of stimuli, each to be assessed on a set of items. The entire measurement design consists of 141 items.

Table 2

The Measurement Design of MSCEIT

Facet	Conditions of facets							
Branch (<i>b</i>)	Perceiving Emotions		Facilitating Thought		Understanding Emotions		Managing Emotions	
Task (<i>t</i>)	Faces	Pictures	Facilitation	Sensations	Changes	Blends	Emotion Management	Emotional Relations
Stimuli (<i>s</i>)	4 faces	6 pictures	5 assignments	5 questions	20 questions	12 questions	5 vignettes	3 vignettes
Items (<i>i</i>)	5 items per face	5 items per picture	3 items per assignment	3 items per question	1 score per question	1 score per question	4 items per vignette	3 items per vignette

From the perspective of Generalizability Theory (GT; Brennan, 1992; Cronbach, Gleser, Nanda, & Rajaratnam, 1972), one may conceive of the branches, tasks, stimuli and items in the design as potential sources of variance. Each potential source of variance constitutes a *facet* in the design, and each facet may consist of one or more *conditions* of measurement. The first branch, Perceiving Emotions, is measured by the two tasks Faces and Pictures. These consist of pictures (stimuli) of faces, abstract art, and landscapes, and one shall rate the extent to which each stimulus expresses various emotions (items). The second branch, Facilitating Thought, is measured by the two tasks Facilitation and Sensations. Each consists of a set of assignments or questions (stimuli), e.g., about to what extent various emotions are useful in performing various assignments, and one shall rate each assignment on a set of items, e.g., particular emotions. The third branch, Understanding Emotions, is measured by the two tasks Changes and Blends and in each task one shall select the most correct answer among five alternatives (items). The fourth and final branch of the model, Managing Emotions, is measured by the two tasks Emotion Management and Emotional Relations, which consists of a set of vignettes (stimuli), e.g., descriptions of a particular situation, where one shall rate the effectiveness of various alternative responses (items).

The measurement design is clearly multifaceted, which means that there exist several potential sources of variance in the design. If one is interested in a total score for the MSCEIT, at least four sources of variance may affect the score. One facet is the branches, a second is the tasks within each branch, a third is the stimuli within each task, and a fourth facet is the items within each stimuli. Variance may be associated with each of these facets, or with interactions between the facets. All facets, or combinations of facets, may be relevant when estimating generalizability (reliability) of scores. In this thesis, only the branch scores will be of interest, not the total score, as the branch scores are usually used in studies with the MSCEIT. The magnitude of each explicit source of variance in the measurement design will be estimated in Paper 1, along with generalizability coefficients for the branch scores.

1.2.3 The Consensus Method Determines Correct Answers

Measuring EI as ability requires item scores that can differentiate between various levels of performance. In the MSCEIT, the correctness of a particular response is defined as the proportion of a larger sample endorsing that response. In *expert scoring*, the sample consists of 21 members of the International Society for Research on Emotions (ISRE), while in general *consensus scoring*, the sample consists of more than 5000 subjects from diverse geographic locations and nations. The scores derived from these two scoring methods tend to correlate to a very high degree (Mayer et al., 2002).

Each item score in the MSCEIT is assumed to reflect a respondent's EI: "For example, if a respondent indicated that surprise was 'definitely present' in a face, and the same alternative was chosen by 45% of the sample, the individual's score would be incremented by the proportion, .45" (Mayer, Salovey, Caruso, & Sitarenios, 2003, p. 100). This means that a person will obtain a higher score on EI the more his or her responses matches the modal responses in the general consensus or expert sample. The modal response

of an easy question will likely have a high consensus value, as most people will detect the correct response. The modal response of a difficult question, however, will likely have a low consensus value as many alternative responses might seem plausible, making it more difficult to choose. Thus, a correct response on a very easy item (e.g., where the consensus value for the modal response is .95) will contribute more to the EI score, and thus reflect more EI, than a correct response on a difficult item (e.g., the consensus value for the modal response is .30). An incorrect response on an easy item, however, will imply a loss of an otherwise large increase in the EI score (e.g., the score might increase only .05 for an incorrect response, instead of .95 for a correct response). A low consensus value for the modal response might not always indicate a difficult item, however. It may also indicate that no response is more correct than other responses, or that all responses are inappropriate. In either case, people's responses will be distributed more or less evenly on the various alternatives.

1.3 The Construct Validity of the Scores from MSCEIT

Because the MSCEIT is the only test that measures the entire four-branch ability model, and because it is frequently used to assess EI, it is important that the scores are valid. Researchers have been concerned with several issues related to the construct validity of scores from MSCEIT (Averill, 2004; Brody, 2004; Conte, 2005; Gohm, 2004; Landy, 2005; Locke, 2005; Matthews, Roberts, & Zeidner, 2004; Mayer, Salovey, & Caruso, 2004a, 2004b; Oatley, 2004; Zeidner, Roberts, & Matthews, 2004). The present thesis will assess the construct validity of the scores from MSCEIT further.

Messick (1995) describes six aspects of construct validity that may be relevant in order to assess the construct validity of scores from the MSCEIT. These six aspects are *content*, *substantive*, *structural*, *generalizability*, *external*, and *consequential* aspects of construct validity. The *content* aspect refers to the extent the selected tasks are representative

of the domain and all important parts of the domain are covered. The *substantive* aspect refers to theoretical rationales for the observed consistencies in test responses. The *structural* aspect refers to whether the structure of scores is consistent with what is known about the construct domain. The *generalizability* aspect of construct validity concerns the extent to which the scores may generalize to the construct domain, e.g., that they are correlated with other tasks representing the same domain. The *external aspect* of construct validity concerns to what extent the scores provide convergent and divergent correlations with external variables. Finally, the *consequential* aspect of construct validity relates to the intended and unintended consequences of score interpretation, an issue that also depends on the validity of the scores.

In Paper 1, the generalizability of scores from MSCEIT will be assessed, and in Paper 2 and Paper 3 the external validity of the scores will be assessed. All six aspects of construct validity will be discussed further in the Discussion.

1.3.1 Assessing MSCEIT Scores from the Perspective of Generalizability Theory

Construct validity entails generalizability, which is related to the concept of reliability. Performance tests of EI have been criticized for providing scores with low reliability (MacCann, Matthews, Zeidner, & Roberts, 2003; Zeidner et al., 2004). While this critique has mainly been aimed toward the MEIS, the MSCEIT has been reported to provide branch scores with adequate reliability. Most studies with the MSCEIT, however, do not report reliability coefficients for the scores in the actual sample of study, but instead the reliability coefficients reported in the User's Manual (e.g., Brackett & Mayer, 2003; Day & Carroll, 2004; Schneider, Lyons, & Williams, 2005).

Only a few studies have estimated reliability coefficients for branch scores in the actual sample of study. In these studies, the reliability coefficients for the branch scores are often substantially lower than those reported in the User's Manual. For instance, for the four

branches Perceiving Emotions, Using Emotions, Understanding Emotions, and Managing Emotions, reliability have been reported to be respectively .88, .61, .75, and .58 (Kafetsios, 2004); .80, .69, .03, and .52 (Zeidner, Shani-Zinovich, Matthews, & Roberts, 2005); .90, .73, .71, and .76 (Palmer, Gignac, Manocha, & Stough, 2005); and .85, .64, .67, .45 (Lopes, Salovey, & Straus, 2003). Thus, in some studies, the reliability coefficients for scores from some of the branches are very low, suggesting that the scores may not generalize well to the intended domain. The estimated reliability coefficients also seem to vary greatly from study to study.

1.3.1.1 Correspondence between Measurement Design and Estimation Procedure

An important issue, which has not yet been addressed in previous studies, is that reliability has been estimated with methods that are not appropriate for the underlying measurement design, which may lead to biased or inflated reliability estimates. That is, the many sources of variance in the measurement design have not been taken into consideration when reliability has been estimated. In previous studies, reliability coefficients have been estimated with formulas based on Classical Test Theory (CTT), such as Cronbach's alpha or split half, using the Spearman-Brown formula (Mayer et al., 2002). Cronbach (2004), however, who developed the alpha formula for over 50 years ago, has stated that the "alpha formula is not strictly appropriate for many tests constructed according to a plan that allocates some fraction of the items to particular topics or processes" (p. 403). For this purpose, Cronbach later developed GT, which will be described later. In the MSCEIT measurement design, the items within each branch may be allocated to several different processes, thus CTT is not appropriate when estimating reliability for scores from this design. For instance, for the branch Perceiving Emotions, the items may first be allocated to one of two tasks, Faces or Pictures. Second, the items in each task may be allocated to one of several stimuli.

The scores are therefore multi-faceted, and “the Spearman-Brown does *not* apply when one generalizes over *more* than one facet” (Brennan, 1992, p. 65). In some cases, “the Spearman-Brown formula predicts a larger value for reliability [...] than that obtained using GT” (Brennan, 2001a, p. 113). Therefore, in Paper 1, an alternative and assumingly more appropriate approach, GT, will be used to estimate generalizability (reliability) coefficients, as GT is a framework for studying the dependability of scores in multifaceted designs (Cronbach et al., 1972). The GT approach will be described further in Section 4.1 and in Paper 1.

1.3.1.2 Relevant Sources of Variance for Estimating Reliability

There is also a need to further investigate the sources of variance in the MSCEIT scores, as some important sources of variance are apparently not accounted for when estimating reliability. That is, several items in the MSCEIT provide scores that correlate weakly or negatively with scores from other items intended to measure the same construct. For this reason, scores from 19 items are excluded from the test before scoring (G. Sitarenios, personal communication, December 11th, 2003). Thus, “the scores are based on 122 out of 141 items because psychometric analyses on the normative sample suggested exclusion of 19 items. These were not deleted from the actual test so as to preserve a balanced layout with the same number of items for all questions” (Lopes et al., 2004, p. 1021). The items from which scores are excluded do not seem to be less representative for the domain than the other items, and there is apparently no reason to exclude them from the scales. This inconsistency may thus suggest that the domain is more complex than assumed or the operationalization of the domain may be questioned. The set of items with negative item-total correlations also tend to vary from sample to sample (see e.g., Lopes et al., 2003, p. 647), suggesting that the operationalization is not optimal.

In GT, one may estimate simultaneously the magnitude of the many sources of variance in the measurement design. This will be done in Paper 1, along with estimation of generalizability coefficients for the branch scores.

1.3.2 Assessing the External Aspect of Construct Validity of MSCEIT Scores

External validity is an important aspect of construct validity, and refers to whether scores relate in expected ways to other theoretically related variables. Some studies have found support for theoretical predictions from MSCEIT scores (Mayer, Caruso et al., 2000; Mayer et al., 2002). Other studies, however, report inconsistent findings. For instance, Puglia, Stough, Carter, and Joseph (2005) found that the scores from MSCEIT could not differentiate between prisoners that were sex offenders and those who were not, though this was expected by theory. In another study, it was found that scores from the Perceiving Emotions branch were unrelated to scores from other more established measures assessing emotion perception ability (Roberts et al., 2006). Though Mayer et al. (2002) state that “we believe that the evidence for the construct validity for the MSCEIT V2.0 is excellent, and that it already surpasses by far that of any other scale in the area of EI” (p. 43), a recent review of the validity evidence draws the opposite conclusion, that there is

“a tendency to selectively report significant results without noting that most of the correlations obtained between the MSCEIT and the dependent variables were nonsignificant...[and]...there is not a single study reported that indicates that EI has nontrivial incremental validity for a socially important outcome variable after controlling for intelligence and personality” (Brody, 2006, p. 179)

Evidently, different researchers come to different conclusions about the construct validity of the MSCEIT scores. More studies are therefore needed on the validity of the scores from MSCEIT.

In this thesis, the external aspect of construct validity for the MSCEIT scores will be assessed in Paper 2 and Paper 3, by relating the scores to variables assumed to be predicted by EI. In Paper 2, the MSCEIT scores will be related to leader empathy and leader effectiveness, while in Paper 3 the scores will be related to ratings of transformational leadership. All these variables may be considered to be strongly related to EI, which will be explained in the following.

1.3.2.1 *Empathy*

There are good reasons to believe that EI should play an important role in empathy. Empathy is “an affective response that stems from the apprehension or comprehension of another’s emotional state or condition, and that is identical or very similar to what the other person is feeling or would be expected to feel” (Eisenberg, 2000, p. 677). Empathy is assumed to be related to EI (Mayer, Caruso et al., 2000; Mayer, DiPaolo, & Salovey, 1990; Mayer & Geher, 1996; Salovey & Mayer, 1990) and “empathy may be a central characteristic of emotionally intelligent behavior” (Salovey & Mayer, 1990, p. 194). All the four abilities outlined in the Mayer-Salovey model of EI appear to be more or less important in empathy. The first ability, Perceiving Emotions, “entails identifying information conveyed by facial expressions, tone of voice, gestures, body posture, color, rhythm, bodily sensations, and other cues. It also involves the capacity to express emotions effectively using such cues” (Lopes et al., 2006, p. 57). This ability may thus enhance empathy by enabling a person to identify which emotions another person is experiencing, and also express these emotions when empathizing with that person. The second ability, Using Emotions, is also likely to be relevant in empathy, as it concerns ability to generate emotions in order to facilitate thought. That is, “this mood generating ability may also play a role in *empathy* – feeling what other people feel” (Caruso & Salovey, 2004, p. 47). The third ability, Understanding Emotions,

encompasses ability to understand emotions. Since comprehending another's emotions is a central aspect of empathy, the ability to understand emotions seems relevant in order to be empathic. "A deep understanding of emotional processes may help one to judge how other people might respond to different situations" (Lopes et al., 2006, p. 57). The fourth and final ability, Managing Emotions, might also be relevant for empathy, as "the ability to manage emotions in self and others encompasses skills involved in [...] empathic listening, and so forth" (Caruso & Salovey, 2004, p. 47). Thus, all four abilities outlined in the Mayer-Salovey ability model of EI seem more or less relevant in empathy.

Several studies have used empathy as a validity criterion for performance measures of EI (Mayer, Caruso et al., 2000; Mayer et al., 1990; Mayer & Geher, 1996). These studies, however, have used *self-reported* empathy as a validity criterion, despite the validity of self-report measures of emotional abilities is questioned. Other people's ratings of empathy, however, may also provide important information about a person's ability to empathize. Paper 2 will therefore relate the scores from MSCEIT also to other ratings of empathy.

1.3.2.2 *Leader Effectiveness*

EI has also been stated to be important for leader effectiveness. The four abilities outlined in the Mayer-Salovey model of EI are "expected to influence people's capacity to interact well with others, communicate effectively, handle conflict, manage stress, perform under pressure, and create a positive work environment...[and]...all these processes are likely to contribute to work performance" (Lopes et al., 2006, p. 55). Thus, when it comes to leadership, one may assume that emotionally intelligent leaders are more effective than other leaders. For instance, emotionally intelligent leaders have been assumed to be more able to develop a collective sense of goals and objectives; instill in others an appreciation of work activities; generate and maintain excitement, enthusiasm, optimism, cooperation and trust;

and encourage flexibility in decision making and change (George, 2000). There is some evidence, mostly unpublished, reporting a positive relationship between MSCEIT scores and leader effectiveness (for a review, see Daus & Ashkanasy, 2005; Lopes et al., 2006). One published study found a positive relationship between MSCEIT scores and leadership after controlling for personality and intelligence (Rosete & Ciarrochi, 2005), while another study found a positive relationship with leader effectiveness only for the scores from Perceiving Emotions and Using Emotions, not for Understanding Emotions and Managing Emotions (Kerr, Garvin, Heaton, & Boyle, 2006). A recent meta-study found an average correlation of .20 between emotion recognition accuracy (the ability purportedly measured by Perceiving Emotions in the MSCEIT) and negotiation effectiveness (Elfenbein, Foo, White, & Tan, 2007). As emotionally intelligent leaders are assumed to be more effective than other leaders, a positive correlation is expected between MSCEIT scores and ratings of leader effectiveness. The relationship between EI and leadership effectiveness will be assessed in Paper 2.

1.3.2.3 Transformational Leadership

In order to assess the validity of scores from MSCEIT further, Paper 3 will relate leaders' scores to ratings of transformational leadership, which should be positively related to EI. Transformational leadership has previously been found to be strongly associated with leader effectiveness and subordinates' well-being (Bass & Riggio, 2006). Among the positive outcomes for subordinates are less stress and higher job satisfaction (Bass & Riggio, 2006); less burnout symptoms (Seltzer, Numeroff, & Bass, 1989); increased trust, commitment, and team efficacy (Arnold, Barling, & Kelloway, 2001); and reduced frustration and increased optimism, which in turn affect performance (McCull Kennedy & Anderson, 2002). EI as ability has been explicitly stated to be important in transformational leadership (Ashkanasy & Tse, 2000; Bass, 2002; Bass & Riggio, 2006; Brown & Moshavi, 2005) and in effective

leadership in general (Caruso, Mayer, & Salovey, 2002; Caruso & Salovey, 2004; George, 2000). Transformational leaders “meet the emotional needs of each employee” (Bass, 1990, p. 21) and “are sensitive to followers’ needs...they show empathy to followers, making them understand how others feel” (Ashkanasy & Tse, 2000, p. 232), and they are assumed to influence subordinates’ emotions in positive ways (for a discussion, see Ashkanasy & Tse, 2000; Sivanathan, Arnold, Turner, & Barling, 2004). It has been stated that “transformational leadership is intrinsically associated with emotional intelligence” (Ashkanasy & Tse, 2000, p. 232) and that “the components of transformational leadership clearly resemble the key components of emotional intelligence” (Ashkanasy, Hartel, & Daus, 2002, p. 325). This suggests that transformational leadership might be an adequate criterion for exploring the validity of the scores derived from the MSCEIT. No known studies have yet been published on the relationship between EI as ability and transformational leadership, and Paper 3 will assess the validity of MSCEIT scores by relating them to rating of transformational leadership.

2 AIMS OF THE THESIS

The aim of this thesis is to assess the construct validity of scores from the MSCEIT. First, Paper 1 will estimate the magnitude of several sources of variance in the MSCEIT measurement design, and estimate generalizability coefficients for scores derived from this test. The external aspect of construct validity of scores will be further assessed in Paper 2 and Paper 3, relating the MSCEIT scores to variables that should be positively related to EI, i.e., empathy and leadership effectiveness (Paper 2) and transformational leadership (Paper 3).

3 METHODS

3.1 *Sample and Procedure*

In Paper 1, the participants were 111 executives (73 men and 38 women, mean age 43) from various businesses in Norway. In Paper 2 and Paper 3, a subsample of these leaders were used ($N = 104$), as some of the leaders did not provide subordinate ratings. All the leaders were volunteers recruited through a Norwegian municipality and the Administrative Research Foundation (AFF) at the Norwegian School of Economics and Business Administration (NHH) in Norway. The participants were leaders in different work organizations in Norway, approximately 70% private and 30% public organizations. In return for participation, they were provided with an extensive feedback on their results on a larger set of questionnaires assessing personality, EI, and leadership behavior. All leaders were asked to select three to five subordinates among those who knew them well, in alphabetical order, to prevent a biased selection. Subordinate ratings of the leaders were used in Paper 2 and Paper 3 and such ratings were obtained for 104 of the leaders (68 men and 36 women). A total of 459 subordinates (53% men and 47% women, mean age 44 years) completed the measures relevant for Paper 2 and Paper 3. Each leader was rated by an average of 4.41 subordinates. The subordinates had worked under their leader for an average of 40 months ($SD = 38$ months), and spent on average 7.3 hours ($SD = 9.4$ hours) a week with their leader, suggesting that they on average were well acquainted with their leader, which is necessary to provide accurate ratings.

3.2 *Measurements*

Several questionnaires were administered to the participants, but only the ones that are reported on in the papers will be described here.

3.2.1 EI

The Norwegian translation of the MSCEIT was used to measure EI. Norwegian consensus scores were used, based on responses from 398 Norwegian respondents. The MSCEIT has been translated by the author of the present thesis and back-translated to English by a bi-lingual psychologist, and accepted by the publisher, MHS, as equivalent to the English version. The Norwegian scores also seem to rank order people very similarly to the International consensus scores, an issue that will be considered in more detail in Paragraph 6.1.4.

3.2.2 Empathy

Two measures of empathy were used; one based on leaders' self-rating and one based on subordinates' ratings of their leader. Self-rated empathy was measured with the empathy scale from BarOn EQ-i (Bar-On, 1997). The BarOn EQ-i is a widely used self-report questionnaire purportedly measuring EI, and the Empathy scale is measuring "*the ability to be aware of, to understand, and to appreciate the feelings of others*" (Bar-On, 1997, p. 16). This scale consists of eight items covering various aspects of empathy, related to comprehending and experiencing other's emotions. Subordinates' ratings of empathy was measured with the empathy scale from BarOn EQ360 (Bar-On, 2002) and the items cover both cognitive and emotional empathy.

3.2.3 Leader Effectiveness

Leader effectiveness was measured by four items previously used in studies of leadership effectiveness (Bass, 1985, p. 205). The questions concern both the effectiveness of the leader and the effectiveness of the leader's work unit. The scores are usually averaged into one leader effectiveness score. Subordinate ratings of leader performance have been

found to correlate with objective measures of performance (Conway, Lombardo, & Sanders, 2001).

3.2.4 Transformational Leadership

Subordinates rated their leader on 16 items from the Multifactor Leadership Questionnaire (MLQ 5X; Bass & Avolio, 2000), designed to measure the four components of transformational leadership behavior, i.e. Idealized Influence, Inspirational Motivation, Intellectual Stimulation, and Individualized Consideration. One sub-dimension of Idealized Influence, relating to subordinates' attributions of leader charisma, was not included in this study, as the focus of the larger study (which this study was a part of) was leader *behavior*, not attributions. Idealized Influence was therefore measured by only one scale, measuring the behavioral aspect of this component.

3.2.5 Personality Traits

The Norwegian translation of the 240-item NEO PI-R (Costa & McCrae, 1992) was used to measure the Five Factor Model (FFM) personality traits, and the responses were scored according to Norwegian norms (Martinsen, Nordvik, & Østbø, 2003).

4 STATISTICAL METHODS

In the present thesis, several statistical methods were used to analyze the scores, methods that are not yet frequently applied in research on EI. In Paper 1, GT (Brennan, 2001a; Cronbach et al., 1972) is used in order to estimate the major sources of variance in the measurement design and to estimate generalizability (reliability) coefficients for the scores. In Paper 2 and Paper 3, Multilevel Confirmatory Factor Analysis (MCFA) with latent variables (Heck & Thomas, 2000; Muthén & Muthén, 1998-2007) is used, as the ratings from subordinates are nested and MCFA takes such dependencies into account.

4.1 **Generalizability Theory**

Because the MSCEIT measurement design is multifaceted, GT is used to estimate generalizability coefficients for the branch scores. As discussed previously, it is important that there is a match between the measurement design and the method used to estimate reliability. CTT, which has been used in previous studies, is not appropriate when the design is multifaceted. GT, however, is developed to estimate generalizability of scores from multifaceted designs (Brennan, 2001a; Cronbach et al., 1972). In GT, one may first use a Generalizability study (G study) to estimate the magnitude of the explicit sources of variance in the design, and then use this information in a decision study (D study) to estimate generalizability coefficients for a particular measurement design of interest. Both these types of studies are used in Paper 1.

First, a G study was used to estimate the relative magnitude of the many explicit sources of variance in the measurement design. These analyses were conducted with the computer program urGenova (Brennan, 2001b), which is developed to estimate variance components for unbalanced designs. The MSCEIT measurement design is unbalanced, as there is an unequal number of measurement conditions in the facets. For instance, in Perceiving Emotions, the two tasks Faces and Pictures are measured by an unequal number of stimuli, i.e., four pictures of faces and six pictures of landscapes and art, respectively. The G study provide variance component that inform us about the relative magnitude of variance reflected in one *single* or *typical* item.

Second, a D study is used to estimate generalizability coefficients for a particular measurement design, using the variance components estimated in the G study as input. As the G study is informative about the relative magnitude of the many sources of variance, this information may be used in order to improve the measurement design optimally. The computer program Genova was used to estimate generalizability coefficients for the present

MSCEIT measurement design, and also some alternative designs. The estimated generalizability coefficient for scores from a particular measurement design is informative about to what extent the scores may generalize to the domain of interest, which in GT terminology is called the universe of generalization.

The GT approach may be used even when the sample size of measurement indicators in each branch (items, stimuli, and tasks) is relatively large compared to the sample size of persons, as is the case in the present study. The person sample in Paper 1 is $N = 111$ and at most, scores from 50 items are included in one single analysis.

The GT approach has seldom been used within the field of research on EI despite many measures in this field have a complex measurement design, where such an estimation method is required in order to obtain more correct estimates. The present study demonstrates how one may estimate reliability of EI measures with multi-faceted measurement designs. In Paper 1, the GT approach is discussed in more detail.

4.2 Multilevel Confirmatory Factor Analysis

In Paper 2 and Paper 3, the scores were analyzed using Multilevel Confirmatory Factor Analysis (MCFA) with latent variables. A multilevel approach is appropriate when data are nested, as is the case when the scores are provided by subordinates rating the same leader. A two-level structural equation model (SEM) approach was therefore used in the analyses (Heck & Thomas, 2000). The model on the within-group level accounts for variance due to differences among subordinates rating the same leader, and the model on the between-group level accounts for variance due to differences among leaders, after the variance due to differences among subordinates rating the same leader is taken into account. Different models may be specified on the within- and between-group levels. The data were analyzed with the computer program Mplus 4.21 (Muthén & Muthén, 1998-2007).

A multi-level latent variable approach is neither used frequently in research on EI, nor in research in organizational psychology generally. For instance, the MLQ 5X is one of the most frequently used measures of transformational leadership, but no known published studies have yet analyzed the scores from this measure within a multi-level approach, despite the data from most published studies are of a multi-level nature. Thus, the present study provides knowledge about the factor structure of this widely used measure of transformational leadership, taking into account the important dependencies in the scores.

5 RESULTS

5.1 *Summary of Paper 1*

The aim of Paper 1 was twofold. First, it was to estimate the magnitude of the various sources of variance in the MSCEIT measurement design, and second, to estimate generalizability coefficients for the branch scores from MSCEIT. A total of 111 executives completed the Norwegian version of the MSCEIT (Mayer et al., 2002).

GT (Cronbach et al., 1972) was used as a framework when analyzing the scores, and the scores were analyzed with the computer softwares urGenova (Brennan, 2001b) and Genova (Crick & Brennan, 1983). Several G studies revealed that the scores reflect considerable amounts of measurement error. Only a small proportion of the G-study variance components are due to the person component, as it explained only 3% - 10% of the variance in the four branches. Some important sources of variance not reported in previous studies with the MSCEIT were identified. In Perceiving Emotions, exploratory and confirmatory factor analyses of the scores with Mplus (Muthén & Muthén, 1998-2007) revealed that the scores were multi-dimensional, not unidimensional as stated in the theory. The distinctions between present and absent emotions, and between positive and negative emotions, were found to be important. The scores from the few items representing present emotions in the stimuli correlated weakly, or negatively, with scores from items representing absent emotions. Scores from these items are regularly excluded by the publisher before scoring the International MSCEIT. In every branch of the MSCEIT, however, some item scores correlate weakly or negatively with the respective task score, and scores from 19 items are regularly excluded by the publisher before scoring the test. As all items were initially selected as conceptually adequate indicators of the domain in question, the inconsistencies may indicate that the constructs are more complex than assumed or that the operationalization is poor.

D studies were then run on the scores from each branch, and generalizability (reliability) coefficients were estimated to be .71, .37, .50, and .46 for the branch scores from Perceiving Emotions, Facilitating Thought, Understanding Emotion, and Managing Emotions, respectively. These coefficients are substantially lower than reported in the User's Manual.

If the magnitude of a generalizability coefficient should be above .50, indicating a larger proportion of true scores variance than error variance, only some of the scores derived from the MSCEIT satisfy this criterion. For the branch Perceiving Emotions, the scores from the task Pictures were not interpretable. For the task Faces, however, three correlated scales were identified, i.e., Factor 1: Ability to identify the absence of positive emotions in mostly sad faces, Factor 2: Ability to identify the absence of particular negative emotions and surprise in mostly sad faces, and Factor 3: Ability to identify absent negative emotions in happy faces. Cronbach's alphas were estimated with SPSS to be .73, .79, and .73 for scores from these three factors, respectively. For the branch Facilitating Thought, only the task Facilitation provided generalizable scores, with an estimated generalizability coefficient of .62. For the branch Understanding Emotions, the overall score was generalizable, though the estimated generalizability coefficient was low, only .50. For the branch Managing Emotions, the scores were not generalizable. The validity of the scores from these scales will be further assessed in Paper 2 and Paper 3.

5.2 Summary of Paper 2

The aim of the study reported in Paper 2 was to assess the validity of the scores that were derived from the MSCEIT in Paper 1, i.e., APE/NE (Absence of Positive Emotions in faces expressing Negative Emotions), ANE/NE (Absence of Negative Emotion in faces that mostly express Negative Emotion), ANE/PE (Absence of Negative Emotion in faces mostly

expressing Positive Emotion), Facilitation and Understanding Emotion. EI is expected to be positively related to both empathy (Mayer, Caruso et al., 2000; Mayer et al., 1990; Mayer & Geher, 1996; Salovey & Mayer, 1990) and leaders effectiveness (e.g., George, 2000). Norwegian leaders (N=104) completed the MSCEIT and a self-report empathy scale (Bar-On, 1997), and were rated by their subordinates (N = 459) on empathy (Bar-On, 2002) and on leader effectiveness (Bass, 1985, p. 205).

Multilevel Confirmatory Factor Analysis (MCFA) was used in the analysis, due to the nested nature of data. The analyses provided a good fit for a model with three factors on the within-group level, representing Subordinate rated Leader Empathy, Leader Effectiveness and Unit Effectiveness, and five factors on the between-group level, representing Self-rated Cognitive Empathy, Self-rated Emotional Empathy, Subordinate-rated Empathy, Leader Effectiveness, and Unit Effectiveness.

The results show that the five EI scales were unrelated to Subordinate-rated Empathy, Self-rated Cognitive Empathy, Unit Effectiveness and Leader Effectiveness. Two of the EI scales, however, APE/NE and ANE/NE, were positively related to Self-rated Emotional Empathy ($r = .30$ and $r = .25$, respectively). Self-rated Cognitive Empathy and Self-rated Emotional Empathy were positively related to Subordinate-rated Empathy, supporting the construct validity of these scales. Overall, only some positive relationships were found between the scores from MSCEIT on the one hand, and empathy and leader effectiveness on the other. The validity of the scores from MSCEIT is therefore questioned.

5.3 Summary of Paper 3

The aim of Paper 3 was to assess the validity of the MSCEIT scores further, by relating the scores to ratings of transformational leadership (Bass, 1997). Transformational leadership is assumed to be positively related to EI (Ashkanasy & Tse, 2000; Bass, 2002;

Bass & Riggio, 2006; Brown & Moshavi, 2005). In this study, leaders (N = 104) completed the MSCEIT (Mayer et al., 2002) and a measure of the FFM personality traits, the NEO PI-R (Costa & McCrae, 1992) and were rated by an average of 4.4 subordinates (N = 459) on transformational leadership behavior, measured by the MLQ 5X (Bass & Avolio, 2000).

Multilevel Confirmatory Factor Analysis (MCFA) was used in the analysis, due to the nested nature of data. The MCFA provided a good fit for a model where the scores from MLQ 5X were represented by four correlated factors on the within-group level (representing Idealized Influence, Inspirational Motivation, Intellectual Stimulation, and Individualized Consideration) and three correlated factors on the between-group level (i.e., Idealized Influence/Individualized Consideration, Inspirational Motivation, and Intellectual Stimulation).

The results show that only one of the derived scales from MSCEIT was related to transformational leadership. That is, the scores from Facilitation was positively related to Inspirational Motivation ($r = .32$, $t = 1.920$). This relationship was not significant, however, after controlling for the personality factors Openness and Agreeableness. As there is a strong theoretical relationship between transformational leadership and EI, these null-findings may question the validity of the scores from the MSCEIT, in line with the findings from Paper 2.

6 GENERAL DISCUSSION

6.1 *The Construct Validity of Scores from MSCEIT*

The aim of the present thesis was to assess the construct validity of scores from the MSCEIT. In Paper 1, the major sources of variance in the MSCEIT design were identified, and generalizability coefficients were estimated for derived scores. In Paper 2 and Paper 3, the external validity was assessed for the scores that were found to be interpretable and generalizable in Paper 1. Together, the three papers suggest that the validity of the scores from MSCEIT may be questioned. Six aspects of the scores' construct validity will be discussed in the following, i.e., content, substantive, structural, generalizability, external, and consequential aspects of construct validity (Messick, 1995).

6.1.1 The Content Aspect of Construct Validity

The content aspect of construct validity refers to the extent to which the selected tasks in a test are representative of the domain to be measured, e.g. to what extent all important parts of the domain are covered, and the technical quality of the scores.

Content validity must be assessed in relation to the domain to be measured. The domains for some of the branches in the four-branch model, however, seem to be defined very broadly (Table 1). For instance, for Perceiving Emotions, the domain encompasses ability to perceive emotions, not only in faces, but also in objects in general, without any restriction of the kinds of objects that are relevant. In the MSCEIT, pictures of stones and landscapes are used to measure this ability, but neither stones nor landscapes are designed to express emotions. It may be that responses from such items to a larger extent reflect individual projections rather than emotion perception ability.

6.1.1.1 *The Items May not Represent all Important Aspects of the Domain*

Content validity requires that tasks, stimuli, and items cover all important aspects of the domain to which the scores are intended to generalize, but one may question to what extent this is the case for the MSCEIT.

For the branch Perceiving Emotions, the items represent both emotions that are present and absent in the stimuli, but the majority of the items are concerned with emotions that are absent. This suggests that the domain for present emotions may not have been sampled adequately. The items also represent positive and negative emotions, and one may in future studies try to sample items and stimuli from this domain more systematically, in order to obtain a more balanced representation of positive and negative emotions, and absent and present emotions.

One may also question whether the items in the branch Facilitating Thought represent the domain for this branch adequately. In the task Sensations, some of the items have an ambiguous meaning. For instance, one of the items asks the respondent to compare the feeling of guilt with e.g., “blue”. The term ‘blue’ has two different meanings in English: it can refer to the *color* blue and the *psychological feeling* of blue, i.e., the feeling of sadness. The authors’ intention, however, is that the term ‘blue’ should be interpreted literally, i.e., as the color blue, not the psychological feeling of “blue” (J. D. Mayer, personal communication, October, 10th, 2001). This interpretation, however, does not seem to be determined by the text. In traditional tests, such ambiguity would result in a low or negative correlation between the item score and scores from other items assessing the same construct (if the other items are appropriate measures of the intended construct). This particular item in MSCEIT, however, is not among those items from which scores are regularly excluded, suggesting that the item does not provide scores that are very different from the other item scores in this task. One

may therefore speculate whether the other items in this task are valid, and whether they represent adequately the domain to which they are to be generalized.

For the branch Understanding Emotions, the items may seem to represent the domain, as they may appear to measure knowledge of emotions. The results from the studies in Paper 1, however, suggest that the items are not very efficient operationalizations of this ability, as one needs a large sample of items to obtain generalizable scores. This may be due to the way the items are scored, with consensus scoring, which will be discussed further later on.

For the branch Managing Emotions, however, one may question to what extent the scores from items, stimuli, and tasks are representative of the domain to which one wants to generalize. The items seem to measure *knowledge* of strategies to manage one's own and others emotions, and not necessarily *ability* to manage one's own and others' emotions. Knowledge of such strategies does not entail that one possess the ability to manage emotions. For instance, a person who has a tendency to experience frequently negative affect might wish to regulate these emotions, and acquire knowledge about this from e.g., self-help books on emotion regulation strategies. Though she may become better able to manage emotions, she might still be less able than an average person to manage own emotions, despite her superior knowledge about such strategies.

6.1.1.2 *The Scoring Method May Be Inappropriate*

An important issue that questions not only the content validity, but all other aspects of validity of scores, is the use of consensus scoring to score item responses. The consensus method implies that easy questions provide large increments in EI, while difficult questions provide small increments in EI. It has been suggested that consensus scoring is more effective in screening for “emotional stupidity” than discriminating between levels of EI at the upper end of the range (Matthews et al., 2004, p. 186). Another weakness with the consensus

method is that it is very sensitive for response biases. Adjacent response alternatives may provide highly different increments in scores, and if a respondent has a tendency to answer all scales somewhat more extreme than what most people do, the person will obtain a very low score on EI. Thus, the consequence of using consensus scoring may be that none of the scores in MSCEIT are valid

As the use of consensus scoring may lead to invalid scores, it may partly explain why an average item score only to a small extent reflect differences between persons and to a large extent reflect measurement error. The items do not seem to be very efficient in measuring individual differences and the results suggest that one may need a very large pool of items in order to obtain generalizable scores. Thus, overall, the content validity of the scores from MSCEIT may be questioned, as the scores from items, stimuli, and tasks in the MSCEIT do not seem to be optimal examples of the domain to which they are meant to represent.

6.1.2 The Substantive Aspect of Construct Validity

The substantive aspect of construct validity concerns the theoretical rationale for observed consistencies in test responses. The theory specifies that the scores from MSCEIT represent four correlated factors, corresponding to the four branches in the model. For instance, Mayer et al. (2003) states that “we believe that the domain of EI is well-described by one-, two-, and four oblique (correlated) factor models” (p. 98). Paper 1, however, revealed some consistencies and inconsistencies that are not accounted for by the theory.

In Perceiving Emotions, there seem to be some consistencies in scores reflecting the distinction between positive and negative emotions, and the distinction between present and absent emotions. The decision to use several items to measure each face, and the inclusion of stimuli other than faces, seem to be based on a need to obtain more reliable scores. That is,

when Mayer, DiPaulo, and Salovey (1990) first developed performance measures of ability to perceive emotions, they criticized the existing measures to include too few items, i.e.,

From a measurement standpoint, the collection of few subject responses or severely restricting participants' responses limits reliability. Measurement also may be improved by considering a wider range of affective stimuli; there are far more emotionally laden stimuli than faces (p. 774).

The results from Paper 1, however, suggest that ability to identify emotions that are not expressed in faces may be different from ability to identify emotions that are expressed. The scores from the few items representing present emotions were weakly, or negatively, correlated to scores from items representing absent emotions, suggesting that ability to identify present and absent emotions might be represented by separate scales. These consistencies are neither accounted for by the theory nor reflected in the present MSCEIT measurement design. Moreover, several inconsistencies seem to be present in scores from the other three branches. The results from Paper 1 reveal that a large proportion of the variance in scores is due to measurement error, or facets not specified in the design. And several items are regularly excluded by the publisher before scoring the International MSCEIT, due to low item-total correlation. No theoretical explanation has been provided for these inconsistencies, and none of the items from which scores are deleted seem to be less conceptually related to the constructs in question. For Faces, the items from which scores are excluded seem to be the ones that are most related conceptually to the construct in question. The set of items with low item-total score correlation also seem to vary from sample to sample, even with the International MSCEIT (see e.g., Lopes et al., 2003, p. 647). Such inconsistencies might be due to poor operationalization, an inadequate scoring method, or more complex constructs

than assumed. As such inconsistencies are not accounted for by the theory, the substantive validity of the scores may be questioned.

6.1.3 The Structural Aspect of Construct Validity

The structural aspects of construct validity entails that the “internal structure of the assessment (i.e., interrelations among the scored aspects of task and subtask performance) should be consistent with what is known about the internal structure of the construct domain” (Messick, 1995, p. 746). Several issues already discussed may be relevant to this aspect of construct validity, such as the consistencies and inconsistencies in scores discussed in the previous paragraph. In the following, however, the evidence for the correspondence between the scores and the theory will be further discussed.

To support the structural validity of scores, the structure in scores should correspond to the theoretical model. Confirmatory Factor Analysis (CFA) is a powerful method to assess such correspondence. The small sample size of persons in the present study, and the relatively large sample of items in MSCEIT, however, did not allow for CFA of the overall structure in scores. The generalizability analyses were therefore conducted under the assumption that the structure of MSCEIT scores corresponds to the model.

Some studies, however, have provided results from CFA indicating a good fit between the scores from MSCEIT and the four-branch ability model (e.g., Day & Carroll, 2004; Mayer et al., 2003). But if one looks closer at how these studies have been conducted, they are not designed to provide optimal information about the four-factor ability model. None of these studies have tested the correspondence between the 141 item scores and the underlying theoretical model. Instead, the 141 item scores have been averaged into eight task scores, or parcels, which are then used in the analysis. The use of parcelling may increase model fit for misspecified models and reduce our ability to identify the model as not fitting the data

(Bandalos & Finney, 2001; Kim & Hagtvet, 2003). Thus, the studies purportedly providing support for the measurement model of MSCEIT are not informative about the structure in the scores, and we do not know whether the scores correspond to the underlying theory. The structure of the item scores may be questioned, however, as several scores correlate weakly or negatively with scores from items assumed to measure the same construct. Also, the set of items providing positive item-total score correlations seem to vary from sample to sample, which might make such fit difficult to obtain. As was demonstrated in Paper 1, the scores from Perceiving Emotions do not constitute one factor, as suggested by theory, but three factors. Thus, as the scores from Perceiving Emotions apparently do not correspond to the theory, and because we do not know whether the scores from the other branches fit with the theory, the structural validity of the MSCEIT scores is highly uncertain. One may therefore question the structural aspect of construct validity of the scores from MSCEIT.

6.1.4 The Generalizability Aspect of Construct Validity

The generalizability aspect of construct validity concerns the “extent to which score properties and interpretations generalize to and across population groups, settings, and tasks (Messick, 1995, p. 745). GT may be used to assess the extent to which scores may generalize to the intended domain. The generalizability coefficients estimated in Paper 1, however, suggest that many of the scores may not generalize to the intended domain. In Paper 1, generalizability was estimated using all items in the MSCEIT, taking into account the explicit sources of variance in the measurement design. The estimated generalizability coefficients for the four branch scores ranged from .46 to .71, which is low, considering that each scale consists of 29 to 50 items. These estimates are also much lower than the reliability estimates provided in the User’s Manual (Mayer et al., 2002). The latter coefficients, however, are likely overestimated, as scores from 19 items with low item-total correlation were excluded

before estimation, and because the applied formulas do not take into account the many sources of variance in the measurement design. As CTT is mainly restricted to one-facet test designs, the application of such estimation procedures on scores from the MSCEIT test design will very likely produce biased reliability estimates. GT, however, provides more relevant information and is an appropriate approach for estimating reliability for scores from the multifacet MSCEIT measurement design. Only some of the scores derived from the present design of MSCEIT did provide scores with estimated generalizability coefficients above .50, which entails that there is a larger proportion of true variance than error variance.

Generalizability also entails the extent to which the scores may generalize to other cultural settings. This issue has not been addressed in this thesis, but the initial G studies suggest that cultural differences, errors in translation, or idiosyncratic properties of the present sample do not have a great influence on the way the scores from MSCEIT rank order people. In fact, the inconsistencies in rank order are smaller between Norwegian and International consensus scores than between International consensus and expert scores. Moreover, most items with low item-total score correlation in the present sample are identical to the items from which scores are regularly excluded in the International MSCEIT due to low item-total score correlation. Thus, the findings in Paper 1 are likely due to properties of the MSCEIT measurement design rather than cultural differences, translation, or particular characteristics of the present sample. As many of the scales from the MSCEIT seem to provide scores with relatively small estimated generalizability coefficients, the scores may not generalize very well to the intended domains, and the generalizability aspect of construct validity is only supported for some of the scales.

6.1.5 The External Aspect of Construct Validity

The external aspect of construct validity concerns to what extent scores relate in expected ways to external variables. This was assessed in Paper 2 and Paper 3.

The finding that the scores were not related to subordinates' ratings of empathy and leader effectiveness is important. EI is assumed to be positively related to both leader effectiveness (Caruso & Salovey, 2004; George, 2000) and empathy (Mayer, Caruso et al., 2000; Mayer et al., 1990; Mayer & Geher, 1996; Salovey & Mayer, 1990). Scores from one of the EI scales was related to one of the transformational leadership factors, but the relationship was not significant after controlling for scores representing the FFM personality dimensions. Previous studies have found that scores from the five personality factors in FFM are predictive of transformational leadership (Bono & Judge, 2004), and several studies report low correlations between scores from MSCEIT and various measures of the FFM (Lopes et al., 2004; Lopes et al., 2003; MacCann et al., 2003). In order for the construct of EI to be useful, it has been argued that measures of EI must predict behavior over and above intelligence and personality traits (Brody, 2004; Ciarrochi, Chan, Caputi, & Roberts, 2001). Thus, the finding that the scores from MSCEIT are unrelated to transformational leadership after controlling for FFM indicate that the scores from MSCEIT do not have incremental validity over the FFM in predicting transformational leadership. Self-rated empathy, however, seems to predict other-rated empathy to some extent, while MSCEIT scores do not. As some of the important predictions from MSCEIT scores were not supported in the present study, the external aspect of construct validity is not supported, and one may question the validity of scores from MSCEIT, as far as the present data is concerned.

When it comes to transformational leadership, at least three unpublished studies referenced in PsychInfo report null-findings between scores from MSCEIT and

transformational leadership. No known studies have yet been published, however, reporting any positive findings between MSCEIT and transformational leadership, even though there is a strong theoretical relation between these constructs.

Previous studies have found that self-rated emotional abilities are largely unrelated to scores from performance measures of these abilities (Brackett et al., 2006). These findings have often been interpreted to mean that self-report is not a valid measure of emotional abilities. The results may, however, also indicate that the MSCEIT scores are invalid. The present study found that self-reported empathy was positively related to subordinates' ratings of empathy, while the MSCEIT scores were mostly unrelated to both of these measures. Such results may therefore suggest that self-reported empathy is a more valid assessment of empathy than the scores from MSCEIT.

The results reported in Paper 2 and Paper 3 did not support the external aspects of construct validity of MSCEIT scores.

6.1.6 The Consequential Aspect of Construct Validity

The consequential aspects of construct validity concern the intended and unintended consequences of score interpretation. This issue has not been assessed in the present thesis. The User's Manual, however, recommends that the MSCEIT may be used in research, corporate, educational, and clinical settings, and in correctional facilities and preventive programs (Mayer et al., 2002). As the validity of the scores may be questioned, however, one should be very cautious of using the MSCEIT in any of these settings, as it may have negative and unintended negative consequences for the people assessed.

6.2 The Scientific Contributions of These Findings

The findings in this thesis are important. The MSCEIT is a widely used measure of EI, developed to measure the four-branch ability model of EI. The concept of EI is relatively

new, and the four-branch model is presently the most elaborate conceptualization of EI, and the MSCEIT is the only measure that provides evidence for the entire model. The present studies suggest that the construct validity of scores from MSCEIT is weak. This either indicates that the domain underlying the MSCEIT is more complex than stated in the EI theory, or that the MSCEIT is a poor operationalization of this domain. The present papers show that the scores do not predict theoretically related variables as empathy, leader effectiveness, and transformational leadership, despite their strong theoretical relationships with EI. When the MSCEIT scores are unrelated to these variables, there is reason to question the validity of the scores.

This thesis also provides some other interesting findings, though not related to the four-branch model of EI. In previous studies, inconsistent results have been reported regarding the factor structure of the scores from the frequently used MLQ 5X. No known studies have yet analyzed the scores within a multi-level framework, despite the score from MLQ 5X are often of a multi-level nature. The present thesis is the first known study to analyze the scores from the MLQ 5X within a multi-level latent variable framework. A good fit was obtained for a model with four correlated factors on the within-group level, and three correlated factors on the between-group level. The present study demonstrates that averaging the scores from transformational leadership into one single score, which is often done, is not supported by the data. Future studies, however, may use MCFA to see whether it is possible to replicate the factor structure found in the present study.

7 CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

The aim of this thesis was to assess the construct validity of scores from MSCEIT and the findings suggest that there are good reasons to question their construct validity. In the future, one might try to improve the measurement design of MSCEIT in ways suggested by

the present studies, in order to provide evidence for the four-branch ability model. At present, however, the items, stimuli, and tasks in MSCEIT do not seem to differentiate well between people, and one may question the validity of the consensus method of scoring the item responses.

One might also try to design an entirely new test of EI, with tasks that measure individual differences in this domain better. One may for example use recordings of interactions among people, as is used in the empathic accuracy approach (Ickes et al., 1997) or in the Interpersonal Perception Task (Costanzo & Archer, 1989). Future studies may also try to design alternative tests that measure the four branch model of EI, in order to see if the data may support the model. The analytical methods used in this thesis may be useful for this purpose. That is, measures of EI are likely to be multifaceted, and the GT approach may be useful to identify the magnitude of the variance associated with various facets in the domain. One may also use the GT approach to optimize the sample sizes of different measurement conditions in order to obtain a desired level of generalizability.

“Validity is an evolving property and validation a continuing process” (Messick, 1995, p. 741) and the present thesis may not definitely rule out that some scales from the MSCEIT may provide valid scores. There are good reasons however, to question the validity of the scores, and one should therefore be very careful in interpreting the scores as indicative of a person’s EI.

8 REFERENCES

- Arnold, K. A., Barling, J., & Kelloway, E. K. (2001). Transformational leadership on the iron cage: which predicts trust, commitment and team efficacy? *Leadership & Organization Development Journal*, 22, 315-320.
- Ashkanasy, N. M., Hartel, C. E. J., & Daus, C. S. (2002). Diversity and emotion: The new frontiers in organizational behavior research. *Journal of Management*, 28, 307-338.
- Ashkanasy, N. M., & Tse, B. (2000). Transformational leadership as management of emotion: A conceptual review. In N. M. Ashkanasy & C. E. Haertel (Eds.), *Emotions in the workplace: Research, theory, and practice*. Westport, CT: Quorum Books/Greenwood Publishing Group.
- Averill, J. R. (2004). A tale of two snarks: Emotional intelligence and emotional creativity compared. *Psychological Inquiry*, 15, 228-233.
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bar-On, R. (1997). *BarOn Emotional Quotient Inventory (EQ-i): Technical manual*. Toronto, Canada: Multi-Health Systems.
- Bar-On, R. (2000). Emotional and social intelligence. Insights from the Emotional Quotient Inventory. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence. Theory, development, assessment, and application at home, school, and in the workplace*. San Francisco: Jossey-Bass.
- Bar-On, R. (2002). *The BarOn EQ-360*. Toronto, Canada: Multi-Health Systems.

- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: The Free Press.
- Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational Dynamics*, 18, 19-31.
- Bass, B. M. (1997). *Transformational leadership: Industrial, military, and educational impact*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bass, B. M. (2002). Cognitive, social, and emotional intelligence of transformational leaders. In R. E. Riggio, S. E. Murphy & F. J. Pirozzolo (Eds.), *Multiple intelligences and leadership* (pp. 105-118). Mahwah, NJ: Lawrence Erlbaum Publishers.
- Bass, B. M., & Avolio, B. (2000). *MLQ: Multifactor Leadership Questionnaire (2nd ed.)*. Redwood City, CA: Mind Garden.
- Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, 89, 901-910.
- Boyatzis, R. E., Goleman, D., & Rhee, K. S. (2000). Clustering competence in emotional intelligence: Insights from the Emotional Competence Inventory. In R. BarOn & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace*. San Francisco: Jossey-Bass.
- Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, 29, 1147-1158.
- Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance

- measures of emotional intelligence. *Journal of Personality and Social Psychology*, *91*, 780-795.
- Brennan, R. L. (1992). *Elements of generalizability theory*. Iowa City, IA: American College Testing Program.
- Brennan, R. L. (2001a). *Generalizability theory*. New York: Springer.
- Brennan, R. L. (2001b). urGenova (Version 2.1) [Computer software]. Available from <http://www.education.uiowa.edu/casma/GenovaPrograms.htm>.
- Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry*, *15*, 234-238.
- Brody, N. (2006). Beyond g. In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* Mahwah, NJ: Lawrence Erlbaum Associates.
- Brown, F. W., & Moshavi, D. (2005). Transformational leadership and emotional intelligence: A potential pathway for an increased understanding of interpersonal influence. *Journal of Organizational Behavior*, *26*, 867-871.
- Byrne, J. C., Dominick, P. G., Smither, J. W., & Reilly, R. R. (2007). Examination of the discriminant, convergent, and criterion-related validity of self-ratings on the Emotional Competence Inventory. *International Journal of Selection and Assessment*, *15*, 341-353.
- Caruso, D. R., Mayer, J. D., & Salovey, P. (2002). Emotional intelligence and emotional leadership. In R. E. Riggio, S. E. Murphy & F. J. Pirozzolo (Eds.), *Multiple intelligences and leadership* (pp. 55-74). Mahwah, NJ: Lawrence Erlbaum Associates.
- Caruso, D. R., & Salovey, P. (2004). *The emotionally intelligent manager*. San Francisco: Jossey Bass.

- Ciarrochi, J., Chan, A., Caputi, P., & Roberts, R. D. (2001). Measuring emotional intelligence. In J. Ciarrochi, J. P. Forgas & J. D. Mayer (Eds.), *Emotional intelligence in everyday life: A scientific inquiry*. Philadelphia: Taylor & Francis.
- Conte, J. M. (2005). A review and critique of emotional intelligence measures. *Journal of Organizational Behavior, 26*, 433-440.
- Conway, J. M., Lombardo, K., & Sanders, K. C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer rating. *Human Performance, 14*, 267-303.
- Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R. Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior, 13*, 225-245.
- Crick, J. E., & Brennan, R. L. (1983). Genova: A generalized analysis of variance system [Computer software]. Available from <http://www.education.uiowa.edu/casma/GenovaPrograms.htm>.
- Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391-418.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurement: Theory of generalizability for scores and profiles*. New York: Wiley.
- Daus, C. S., & Ashkanasy, N. M. (2005). The case for the ability-based model of emotional intelligence in organizational behavior. *Journal of Organizational Behavior, 26*, 453-466.

- 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 and Individual Differences, 36*, 1443-1458.
- Eisenberg, N. (2000). Empathy and sympathy. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd ed.). New York: The Guilford Press.
- Elfenbein, H. A., Foo, M. D., White, J., & Tan, H. H. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior, 31*, 205-223.
- George, J. M. (2000). Emotions and leadership: The role of emotional intelligence. *Human Relations, 53*, 1027-1055.
- Gohm, C. L. (2004). Moving forward with emotional intelligence. *Psychological Inquiry, 15*, 223-227.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- Goleman, D. (1998). What makes a leader? *Harvard Business Review* (November-December), 92-102.
- Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ickes, W., Marangoni, C., & Garcia, S. (1997). Studying empathic accuracy in a clinically relevant context. In W. Ickes (Ed.), *Empathic accuracy* (pp. 292-310). New York: The Guilford Press.
- Kafetsios, K. (2004). Attachment and emotional intelligence abilities across the life course. *Personality and Individual Differences, 37*, 129-145.
- Kerr, R., Garvin, J., Heaton, N., & Boyle, E. (2006). Emotional intelligence and leadership effectiveness. *Leadership & Organization Development Journal, 27*, 265-279.

- Kim, S., & Hagtvet, K. A. (2003). The impact of misspecified item parceling on representing latent variables in covariance structure modeling: A simulation study. *Structural Equation Modeling, 10*, 101-127.
- Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *Journal of Organizational Behavior, 26*, 411-424.
- Locke, E. A. (2005). Why emotional intelligence is an invalid concept. *Journal of Organizational Behavior, 26*, 425-431.
- Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I., & Salovey, P. (2004). Emotional intelligence and social interaction. *Personality and Social Psychology Bulletin, 30*, 1018-1034.
- Lopes, P. N., Côté, S., & Salovey, P. (2006). An ability model of emotional intelligence: Implications for assessment and training. In V. U. Druskat, F. Sala & G. Mount (Eds.), *Linking emotional intelligence and performance at work: Current research evidence with individuals and groups*. Mahwah, NJ: Lawrence Earlbaum Associates.
- Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences, 35*, 641-659.
- MacCann, C., Matthews, G., Zeidner, M., & Roberts, R. D. (2003). Psychological assessment of emotional intelligence: A review of self-report and performance-based testing. *International Journal of Organizational Analysis, 11*, 247-274.
- Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology, 68*, 854-870.
- Martinsen, Ø., Nordvik, H., & Østbø, L. E. (2003). *Norsk utgave av Revised NEO Personality Inventory (NEO PI-R) [Norwegian version of the Revised NEO Personality Inventory (NEO PI-R)]*. Oslo, Norway: Gyldendal Akademisk.

- 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., Roberts, R. D., & Zeidner, M. (2004). Seven myths about emotional intelligence. *Psychological Inquiry, 15*, 179-196.
- Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and myth*. Cambridge, MA: The MIT Press.
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298.
- Mayer, J. D., DiPaolo, M., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment, 54*, 772-781.
- Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22*, 89-114.
- Mayer, J. D., Roberts, R. D., & Barsade, S. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59*, 507-536.
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications*. New York: Basic Books.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2000). Models of emotional intelligence. In R. Sternberg (Ed.), *Handbook of Intelligence* (pp. 396-420). New York: Cambridge University Press.

- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). User's manual*. Toronto, Canada: Multi-Health Systems.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2004a). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15*, 197-215.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2004b). A further consideration of the issues of emotional intelligence. *Psychological Inquiry, 15*, 249-255.
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.
- McColl Kennedy, J. R., & Anderson, R. D. (2002). Impact of leadership style and emotions on subordinate performance. *Leadership Quarterly, 13*, 545-559.
- 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.
- Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (Fourth ed.). Los Angeles, CA: Muthén & Muthén.
- Nowicki, S., & Duke, M. P. (2001). Nonverbal receptivity: The diagnostic analysis of nonverbal accuracy (DANVA). In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 183-198). Mahwah, NJ: Lawrence Erlbaum Associates.
- Oatley, K. (2004). Emotional intelligence and the intelligences of emotions. *Psychological Inquiry, 15*, 216-223.
- Palmer, B. R., 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.

- Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525-554.
- Puglia, M. L., Stough, C., Carter, J. D., & Joseph, M. (2005). The emotional intelligence of adult sex offenders: ability based EI assessment. *Journal of Sexual Aggression, 11*, 249-258.
- Roberts, R. D., Schulze, R., O'Brien, K., MacCann, C., Reid, J., & Maul, A. (2006). Exploring the validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) with established emotions measures. *Emotion, 6*, 663-669.
- Rosete, D., & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadership and Organization Development Journal, 26*, 388-399.
- Sala, F. (2006). The international business case: Emotional intelligence competencies and important business outcomes. In V. U. Druskat, F. Sala & G. Mount (Eds.), *Linking emotional intelligence and performance at work. Current research evidence with individuals and groups*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9*, 185-211.
- Salovey, P., & Mayer, J. D. (2002). The positive psychology of emotional intelligence. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 159-171). New York: Oxford University Press.
- Schmidt, F. L., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

- Schneider, T. R., Lyons, J. B., & Williams, M. (2005). Emotional intelligence and autonomic self-perception: Emotional abilities are related to visceral acuity. *Personality and Individual Differences, 39*, 853-861.
- Seltzer, J., Numeroff, R. F., & Bass, B. M. (1989). Transformational leadership: Is it a source of more or less burnout or stress? *Journal of Health and Human Resource Administration, 12*, 174-185.
- Sivanathan, N., Arnold, K. A., Turner, N., & Barling, J. (2004). Leading well: Transformational leadership and well-being. In P. A. Linley & S. Joseph (Eds.), *Positive psychology in practice* (pp. 241-255). Hoboken, NJ: John Wiley and Sons.
- Sternberg, R., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive validity of IQ. *Merrill-Palmer Quarterly, 47*, 1-41.
- Zeidner, M., Roberts, R. D., & Matthews, G. (2004). The emotional intelligence bandwagon: Too fast to live, too young to die? *Psychological Inquiry, 15*, 239-248.
- Zeidner, M., Shani-Zinovich, I., Matthews, G., & Roberts, R. D. (2005). Assessing emotional intelligence in gifted and non-gifted high school students: Outcomes depend on the measure. *Intelligence, 33*, 369-391.

9 PAPERS I – III