Theory and Measurement of Social Intelligence as a Cognitive Performance Construct

Dissertation

zur Erlangung des akademischen Grades

Doktor der Philosophie,

genehmigt durch die Fakultät

für Geistes-, Sozial- und Erziehungswissenschaften

der Otto-von-Guericke-Universität Magdeburg

von Dipl.-Psych. Susanne Weis

geb. am 27.08.1974 in Frankenthal / Pfalz

Gutachterin/Gutachter: Prof. Dr. Heinz-Martin Süß

Gutachterin/ Gutachter: Prof. Dr. Paul Probst

Eingereicht am: 22.08.2007

Verteidigung der Dissertation am: 21.01.2008

Acknowledgements

First and foremost, I want to express my gratitude to my supervisor and mentor Heinz-Martin Süß for many precious years of cooperation and support. I particularly appreciate his elaborate and sophisticated criticisms during the course of the present work and his unparalleled way of maintaining high-quality standards. My special thanks go to Paul Probst who introduced social intelligence in the German empirical research literature, for his willingness to review my dissertation. I would like to thank my colleague and fellow combatant Kristin Seidel for the unprecedented cooperation throughout the course of the project which will hopefully continue on into the future. I enjoyed a lot our vivid discussions, the openness of sharing any thoughts and feelings, and the joy of working together. My thanks go to the staff of our project and the department, in particular to our research assistants, interns, and diploma students Jenny Papenbrock, Janet Feigenspan, Saskia Sage, Alexandra Brett, Kerstin Bremer, Anika Fischer, Ulrike Richter, Eva Wohner, Janine Nötzold, Nadine Hähnel, and Christiane Karthaus. Without their help and inestimable engagement, this work could not have been carried out. Christiane, thanks so much for your gripe, and especially, for just being there during the final countdown. My special thanks go to Birgit Müller as the heart of our team for her proactive and supportive work. I would like to thank my friends and family who, as target persons, allowed insight into their lives through our camcorders, voice recorders, and photo cameras. This work could not have been carried out without their willingness to participate and to share their intimate experiences, thoughts, and feelings with us. I would also like to thank Jennifer Minsky and Namrata Tognatta for showing patience and persistence in proof reading and providing precious comments that contributed to the present work.

I also want to express thanks to the Deutsche Forschungsgemeinschaft (DFG) which funded the project “Kognitive Facetten Sozialer Intelligenz” (Grant SU 196/3-1) between November 2003 and September 2006.

Last but not least, I want to express my gratitude to my family for their never ending support. Particularly, I want to thank my parents for providing me with the opportunity to study, for being committed to my interests and for promoting intellectual engagement and scientific work by teaching me critical thinking and optimism. An inexpressable “Thank you” goes to Lars for being tolerant, for deliberately diminishing his own importance for promoting my interests, and for covering my back from private obligations. Ultimately, thanks to Lars and Sina for keeping me down-to-earth and showing me the truly important things in life.

TABLE OF CONTENTS

<u>INDEX OF TABLES AND FIGURES</u>

Tables

Figures

INDEX OF ABBREVIATIONS AND TEST LABELS

BIS-C	BIS-Creativity ability domain
BIS-F	BIS-Figural-Spatial ability domain
BIS-M	BIS-Memory ability domain
BIS-Model	Berlin Intelligence Structure Model
BIS-N	BIS-Numerical ability domain
BIS-R	BIS-Reasoning ability domain
BIS-S	BIS-Processing Speed ability domain
BIS-Test	Berlin Intelligence Structure Test
BIS-V	BIS-Verbal ability domain
CARAT	Communication of Affect Receiving Ability
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CI	Confidence interval
CRT	Choice Reaction Time Task
DANVA2	Diagnostic Analysis of Nonverbal Accuracy Scale
DF	Degrees of freedom
DFG	Deutsche Forschungsgemeinschaft
DVD	Digital Video Disk
EARS	Emotion Accuracy Research Scale
EC	Emotional competence
ECS	Expert consensus scoring
EI	Emotional intelligence
FEEL	Facially Expressed Emotion Labeling
FPI-R	Freiburg Persönlichkeitsinventar – revidierte Auflage 2001
g_c	Crystallized intelligence
GCS	Group consensus scoring
g_f	Fluid intelligence
GVEESS	Geneva Vocal Emotion Expression Stimulus Set
GWSIT	George Washington Social Intelligence Test
IIP-C	Inventory of Interpersonal Problems – Circumplex
IPT-15	Interpersonal Perception Task – 15
ICI	Interpersonal Competence Inventory
JACBART	Japanese and Caucasian Brief Affect Recognition Test
LEAS	Level of Emotional Awareness Scale
LMT	Long Term Memory Task
MEIS	Multifactor Emotional Intelligence Scale
MERT	Multimodal Emotion Recognition Test
MSCEIT	Mayer-Salovey-Caruso Emotional Intelligence Test
MT	Mouse Speed Task

MTMM	Multitrait-multimethod
NEO FFI	NEO Five Factor Inventory
PC	Personal computer
PONS	Profile of Nonverbal Sensitivity
RMSEA	Root Mean Square Error of Approximation
RT	Reaction time
S1	Study 1
S2	Study 2
SB	Social Behavior Questionnaire
SBS	Standards-based scoring
SC	Social competence
SCF-P	Pictorial Social-Cognitive Flexibility
SCF-V	Verbal Social-Cognitive Flexibility
SEIS	Schutte Emotional Intelligence Scale
SI	Social Intelligence
SIM	Social Intelligence Test Magdeburg
SIT	Chapin Social Insight Test
SJT	Situational Judgment Test
SM	Social Memory
SMa1	Memory for Conversations
SMa2	Memory for Voices
SMf1+2	Memory for Situations – Videos
SMp1	Memory for Couples
SMp2	Memory for Situations - Pictures
SMv1+2	Memory for Written Correspondence
SOI	Structure of Intellect Model
SP	Social Perception
SPa1	Perception of Social Cues in Spoken Language
SPa2	Perception of Emotions in Voices
SPf1	Person Perception – Videos
SPf2	Perception of Body Language – Videos
SPp1	Person Perception – Pictures
SPp2	Perception of Body Language – Pictures
SPv1	Perception of Social Cues - Texts
SPv2	Perception of Body Language – Pictures
SPSS	Statistical Package for Social Sciences
SRMR	Standardized Root Mean Square Residual
SRT	Simple Reaction Time Task
SSI	Social Skills Inventory
STEM	Situational Test of Emotion Management
STEU	Situational Test of Emotional Understanding
SU	Social Understanding
SU_BS	Scenario of target Bringfried
SU_CK	Scenario of target Conny
SU_CP	Scenario of target Christoph

SU_FB	Scenario of target Friedrich
SU_HR	Scenario of target Hannah
SU_KL	Scenario of target Katharina
SU_MM	Scenario of target Matthias
SU_RF	Scenario of target Renate

SUps	Social Understanding scale based on personality ratings
TEMINT	Test of Emotional Intelligence
TKIM	Tacit Knowledge Inventory for Managers
TS	Target scoring
Vocal-I	Vocal Emotion Recognition Test
WAIS-R	Wechsler Adult Intelligence Scale – Revised Edition
WMC 0.18	Experimental Software WMC 0.18

1 Summary

The aim of the present work was to advance research on social intelligence and to establish it as a viable new ability construct within the nomological network of existing human ability constructs. The present work adhered to the requirements on a new intelligence construct as established in the research literature on intelligence (Matthews, Zeidner, & Roberts, 2005; O'Sullivan, 1983; Schaie, 2001; Süß, 2001; Weber & Westmeyer, 2001). In accordance with these requirements, the focus of the present work firstly, was on the theoretical and methodological substantiation of the construct, and secondly on the development of a test battery of social intelligence (i.e., the Social Intelligence Test Magdeburg, SIM. The most central research questions concerned the investigation of the psychometric properties of the newly developed tasks and the construct validity of social intelligence as assessed by the SIM.

Theoretical Foundations

Social intelligence, in the present study, is based on the performance model of Weis and Süß (2005; see also Weis, Seidel, & Süß, 2006) and only contains cognitive ability requirements. The model originally represented a structural model distinguishing between social understanding, social memory, social perception, and social creativity as the cognitive ability domains. The model is modified in the present work by adding a hierarchical assumption in terms of a higher-order social intelligence factor. In extension to this differentiation of operative requirements, further taxonomic considerations are applied which identify additional relevant classificatory principles in definitions of social intelligence: *(a)* contents or cues (e.g., the tone of voice in spoken language or gestures displayed in pictures or videos), *(b)* the queried modalities (i.e., the requested output of a task such as the judgment of emotions or personality traits), *(c)* the settings (i.e., the surrounding contexts), and *(d)* the targets (i.e., as the objects of the cognitive operation such as the self or others, familiar persons or strangers). These taxonomic considerations do not claim to be exhaustive, rather they serve as classificatory principles of existing measurement approaches and as foundations for the subsequent test development.

In the context of the theoretical foundation, related constructs such as emotional intelligence are also addressed where the most central theoretical models are described and discussed. Substantial theoretical shortcomings are identified, related to the current state of conceptualization of the emotional intelligence construct (i.e., the Four-Branch-Model of

Emotional Intelligence; Mayer & Salovey, 1997). These concern the differentiation of emotional processes and contents, and cognitive processes and contents, and the postulated performance requirements of the Four Branches. It is mainly criticized that the interplay of emotional and cognitive processes is not clarified by the Four-Branch-Model and that some Branches contain behavioral and knowledge requirements. For the lack of comprehensive empirical evidence, the present work elaborates the construct overlap of social and emotional intelligence relying on theoretical accounts. Furthermore, domain-specific overlap is identified.

Methodological Foundations

Prior to the description of the test development, existing measurement approaches and the surrounding methodological problems are described with a strong focus on cognitive ability tests. In this context, the problems of item origin, the relevance of context information, different response formats, and scoring procedures are discussed. Against this background and referring to the preceding taxonomic considerations, existing measurement approaches to assess social intelligence as a cognitive performance construct are sampled and discussed in light of the methodological shortcomings and the resulting validity evidence. The conclusions from these considerations identify problems surrounding the use of artificial and decontextualized item material, the use of only written language contents, the adequate scoring procedure, and a mismatch between the purported measurement construct and the actual task requirements. As a consequence, many existing measurement approaches lack evidence for the convergent and divergent construct validity.

Test Development and Research Questions

The test development in the present work was based on the performance model of Weis and Süß (2005) and the associated taxonomic considerations. The test design cross-classified three operative ability domains (i.e., social understanding, social memory, and social perception) and four material related content domains (i.e., written and spoken language, pictures, and videos). This classification resulted in a 3 x 4 multitrait-multimethod design and was foremost intended to balance method-related variance. Additionally, all tasks systematically varied the type of setting (i.e., private vs. public) and the number of persons involved in the situations (i.e., one person, a dyad, and small groups). All tasks relied on genuine task material that was sampled in natural settings involving real persons.

The test principles of the three operational ability domains will be summarized briefly in the upcoming sections. *(a)* The *social understanding* tasks relied on a scenario approach applying the so-called postdiction paradigm. The tasks consisted of eight scenarios, each related to one target person. One scenario consisted of a specific number of scenes which were related to the four material related content domains. Based on each of the scenes, subjects were required to judge the emotions and the cognitions of the target persons, and the relationship of the target to the other persons displayed in the scene. At the end of each scenario, subjects had to rate the personality traits of the target persons and answer control questions about the perceived sympathy, the assumed similarity, and the self-assessed empathic compassion and the perspective taking skills for the respective target person. The response format in the final task version was a 7-point rating scale. Answers were scored in terms of the weighted difference to the target answer (i.e., target scoring). *(b)* The tasks of *social memory* required subjects to watch, read or listen to extracts from social situations. They had to answer multiple-choice or open-ended questions about socially relevant details. Presentation and answering times were limited, and answers were scored in terms of the proportion of correct answers. *(c)* The test principle of the *social perception* tasks required the quick identification of previously presented targets within social situations (i.e., within a written or spoken sentence, within a picture or a video). Targets were, for example, the utterance of an emotion, a disagreement, a specific name, a person wearing different types of clothes or more interactive social facts such as touch or eye contact. Subjects had to strike a key as soon as they detected the target. The answering format was the response latency between the stimulus presentation and the key stroke.

The pivotal research questions concerned the investigation of the psychometric properties of the social intelligence tasks and the examination of the construct validity of social intelligence as assessed by the SIM. The hypotheses of the construct validity established testable models of *(a)* the internal structure of social intelligence as postulated in the design of the test battery and *(b)* the relationship between the broad ability factors of social and academic intelligence (divergent construct validity). Additionally, it was expected that social intelligence would show divergent construct validity with personality traits. Amongst others, some further research questions concerned the convergent construct validity with a measure of nonverbal sensitivity (i.e., a construct related to social and emotional abilities), the exploration of gender differences, the exploration of different scoring methods for the social understanding tasks, and the investigation of the faceted structure of social understanding.

Method: Studies and Materials

The present work was based on two main studies. One hundred twenty six German university students participated in Study 1. The mean age was 21.35 (*sd* = 3.06), and 53.5 % were females. In Study 2, an unselected sample of adults was applied. Participants were between 23 and 40 years old (m_{age} = 28.69; *sd* = 5.57) and 58.8 % of the subjects were female. The sample consisted of heterogeneous subjects in terms of age, education, and occupation.

Both studies applied the test battery of social intelligence and the Berlin Intelligence Structure Test (BIS-Test; Jäger, Süß, & Beauducel, 1997) as a measure of academic intelligence. Additionally, several trait inventories of personality were applied. In both studies, additional measures were utilized that were not directly related to the research questions, thus they are not mentioned at this point.

Empirical Results and Discussion

In addition to the pivotal research questions, Study 1 was particularly directed at the investigation of adequate item and response formats and of the presentation and answering times of the social intelligence tasks. However, the present summary addresses both studies without detailing the concrete steps of test development in between. The focus is rather on the answers to the research questions.

Referring to the psychometric properties, most of the newly developed tasks showed sufficient reliability coefficients. In Study 2, Cronbach's alpha of the social understanding tasks ranged between .75 and .85. The reliability coefficients of the social memory tasks showed a large range (between .19 and .84) with rather low coefficients for one auditory and two pictorial tasks. The reliabilities of the social perception tasks were the largest and ranged between .71 and .98. After some necessary steps of data cleaning (i.e., dealing with missing values, correction of outliers, and trimming of the reaction time scores), the final scales were found to be normally distributed.

With respect to the internal structure of social intelligence, confirmatory factor analysis supported a two-factor structural model with two correlated factors of social understanding and memory. The factor intercorrelations were $r = .35$ in Study 1 and $r = .20$ in Study 2. No higher-order general social intelligence factor was supported in Study 2. Results also showed good data fit of a faceted model of social intelligence with two correlated operative (i.e., social understanding and memory) and two correlated content related factors (i.e., language-based and language-free contents). The factor intercorrelations were $r = .25$

and .26, respectively for the operative and the content factors. However, the loadings on both content related factors were rather heterogeneous, rendering the factors difficult to interpret.

Regarding the construct validity, confirmatory factor analysis supported discriminable social and academic intelligence factors with low correlations between social understanding and the BIS-Reasoning factor (r = .00 in Study 1 and r = .14 in Study 2). The social memory factor and BIS-Memory were substantially correlated with r = .42 in Study 1 and r = .67 in Study 2. However, the social intelligence structural model proved structure independency from academic intelligence: when BIS-Test variance was partialled out of the single tasks, confirmatory factor analysis replicated the structural model of social intelligence based on the residuals. Furthermore, correlational analysis supported the divergent construct validity of the social intelligence tasks with personality traits. The correlational analysis in Study 2 did not support the convergent construct validity with a measure of nonverbal sensitive.

The exploration of gender effects showed only a few significant gender differences in the tasks of both studies, most of them in favor of women. Surprisingly, males tended to perform better on the social understanding tasks in Study 2 while females performed better on the tasks in Study 1. However, both effects were not significant. An exploration of alternative scoring methods for the social understanding tasks revealed substantial weaknesses of group consensus and correlations-based scoring methods. Target and group consensus scoring of the respective content related scales were highly intercorrelated with a smaller correlation between the written language scales (r = .314 - .783). However, Study 2 showed that correlations between group consensus and target scoring were difficult to interpret. The correlation size depended on the item difficulty. Moreover, it was demonstrated that the bivariate distributions between target and group consensus scoring revealed a curvilinear relationship under certain conditions. With respect to correlations-based scoring methods, Study 2 showed very low reliabilities and comparably low intercorrelations between the respective content related scales (r = .271 - .430).

Regarding the faceted structure of the social understanding tasks, these were based on a 4 x 3 x 2 design cross-classifying four content domains (i.e., written and spoken language, pictures, and videos), three modality domains (i.e., emotions, cognition, and relationships), and two setting domains (i.e., private and public). Confirmatory factor analysis supported the modality facet with one factor related to the judgment of emotions and cognitions, and one factor related to the judgment of relationships. Analysis further supported the differentiation

into a language-based and a language-free factor. However, the loadings were heterogeneous on the content factors, making the factors difficult to interpret.

Besides the discussion of the aforementioned research results and their consequences for the social intelligence construct, the test approach and the underlying taxonomic considerations are discussed. Several decisions during the course of test development and the resultant tasks are subject to critical discussion such as the application of genuine task material, the target scoring procedure, the process of item construction and selection, and the cognitive processes underlying the social understanding tasks. Finally, some considerations about test modifications and extensions are addressed. Complementary interesting research questions that add to the current results are also discussed.

2 Introduction

Research on human abilities is more than 100 years old (Spearman, 1904). Research on social intelligence started only a few years after Spearman (1904) introduced academic intelligence (Thorndike, 1920). Thus, social intelligence was one of the first candidates for a new intelligence construct to complement traditional human ability concepts. In a series of several researchers, Landy (2005, 2006) was the last to outline the history of social intelligence, while Walker and Foley (1973), Orlik (1978), Matthews, Zeidner, and Roberts (2002), and Weis and Süß (2005; see also Weis et al., 2006) have all reviewed the literature in terms of theoretical and empirical accounts. In recognition of its long although disputed history, Landy (2006) referred to research on SI as a long, frustrating, and fruitless search, presently ending up in its "replacement with the more modern term emotional intelligence" (p. 81).

Today, emotional intelligence represents a younger candidate for a new ability construct. It was introduced by Salovey and Mayer in 1990. Subsequent attempts to establish emotional intelligence as a new intelligence construct were faced with still ongoing controversial discussions and its utility is still questioned today (Weber & Westmeyer, 2001; Zeidner, Matthews, & Roberts, 2001; Landy, 2006; Ashkanasy & Daus, 2005; Mayer, Salovey, & Caruso, 2000; see also a discussion forum in the German journal '*Zeitschrift für Personalpsychologie'* in 2002, volume 3 with contributions from Asendorpf, Heller, Neubauer & Freudenthaler, Schmidt-Atzert, and Schuler). Thus, why did Landy speak of the replacement of social by emotional intelligence? Why can nearly every single one of the aforementioned reviews on social intelligence be found in the context of publications on emotional intelligence? And why did Goleman entitle social intelligence as "The new science of human relationships" in his recently published book (2006). An answer to these questions would surely require a longer examination than there is space in this introduction. Hopefully, some answers will be provided throughout the present work.

At present, this controversially discussed development of social and emotional ability constructs should be illustrated in more detail. It is, for example, also reflected in the number of publications in scientific search engines (see Fig. 2.1 for the number of publications in PsychInfo). Publications before 1960 are not displayed. Before 1960, emotional intelligence (or the term emotional competence) was not apparent. The number of publications ranged from two to 37 (with a peak around the early 1930s) for social intelligence and from zero to 28 for social competence.

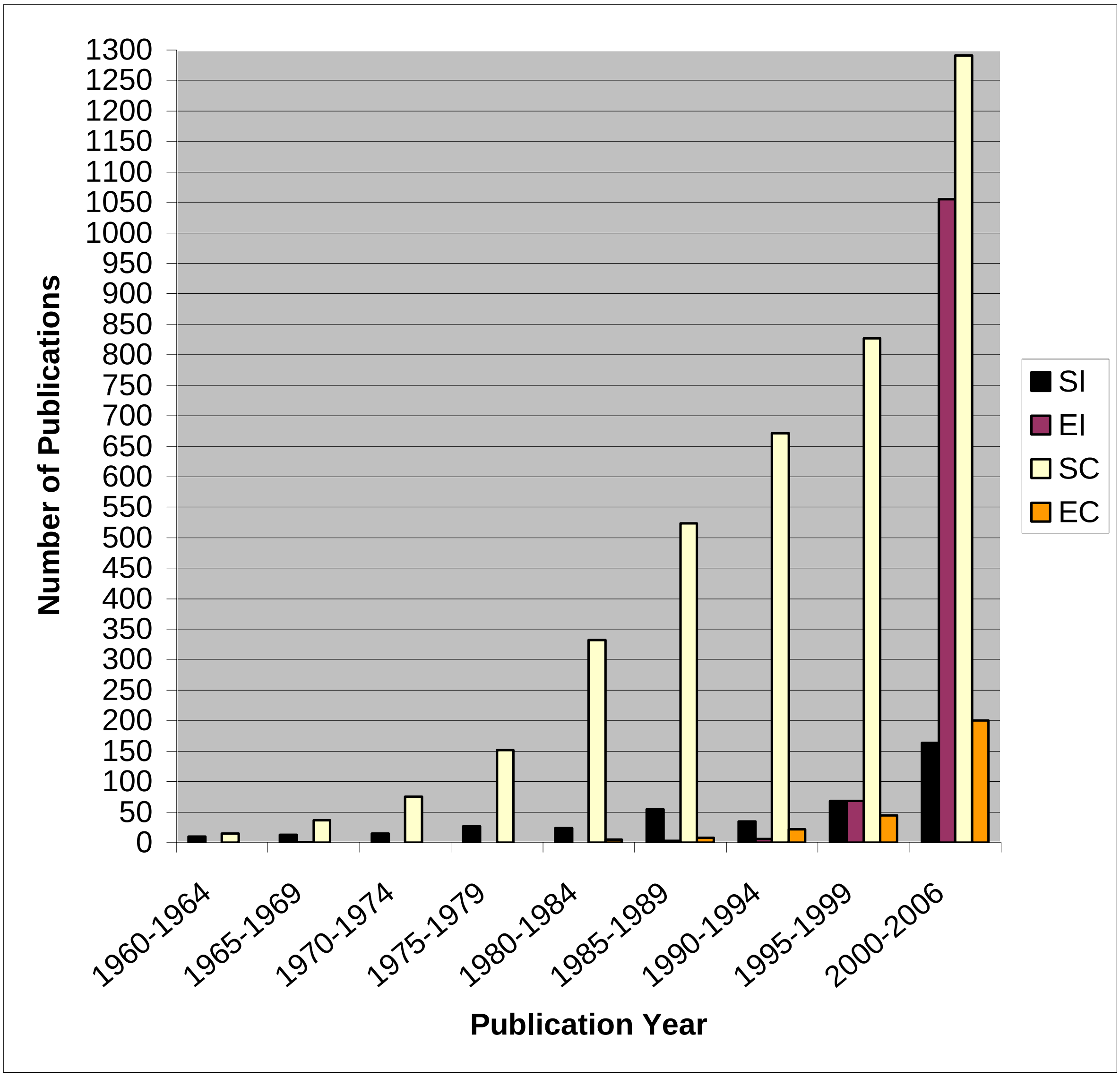


Figure 2.1

Number of Publications of Social and Emotional Intelligence, Social and Emotional Competence

Note. SI = social intelligence, EI = emotional intelligence, SC = social competence, EC = emotional competence.

From its introduction in 1990, the number of publications on emotional intelligence has exploded from 6 between 1990 and 1994 up to 1055 between 2000 and 2006. Social intelligence could not exhibit an approximately comparable career. Moreover, 16 % of the publications on social intelligence appeared in the context of publications on emotional intelligence since the introduction of the latter in 1990.

The diverse careers of intelligence constructs, including academic, social, and emotional intelligence as human ability constructs can be attributed to several reasons. Some of them will be shortly mentioned hereafter and discussed in detail throughout this thesis. Academic intelligence represents the prototype of an intelligence construct. Consequently, every newly emerging candidate for an intelligence construct has to compete with the standard set by academic intelligence. Social intelligence was originally defined as "the ability to understand and manage men and women, boys and girls, and to act wisely in human relations" and was distinguished from abstract and mechanical intelligence (Thorndike, 1920, p. 228). Thorndike (1920) conceived social and mechanical intelligence as logical complements to academic intelligence. He required "a genuine situation with real persons" for the measurement of social intelligence (p. 231). In his conception, social intelligence consisted of cognitive and behavioral requirements. However, from the outset, there were some major problems encountered during the establishment as an ability construct, which exists even today. One of the most striking problems represents the lack of genuineness of task material and task requirements. Assessment approaches were equivalent to those of academic intelligence tests; it was rare that the behavior of genuine persons served as stimuli. Instead, task material consisted of verbal descriptions of social situations or black and white paintings. Social intelligence was not separated conceptually and empirically from academic intelligence so that the general value of social intelligence as a viable intelligence construct was questioned.

Emotional intelligence demonstrated an exponential gain of scientific attention (see Figure 2.1). However, the research has been criticized for committing some substantial errors, some of which are comparable to those found in social intelligence research. Some authors criticize emotional intelligence for applying a labeling approach which renames already existing constructs (e.g., alexithymia, emotional regulation, appraisal of emotions, or even social intelligence) into emotional intelligence (Hedlund & Sternberg, 2000). Others strongly question the value and justification of emotional intelligence as an intelligence construct (Asendorpf, 2002; O'Sullivan, 2007; Weber & Westmeyer, 2001). Thus, how could emotional intelligence attract such large numbers of researchers to publish 1.055 peer-reviewed articles, book chapters, books, or dissertations between 2000 and 2006? To some extent, this can be attributed to the attention emotional intelligence received after the publication of Goleman's book *“Emotional Intelligence – Why it can matter more than IQ”* (1995). After emotional intelligence was introduced by Salovey and Mayer (1990), only six publications emerged until 1994. After Goleman’s book in 1995, the number of publications rose up to 68 within the next five years which meant an 11.3 times multiplication factor. The practical significance of constructs such as social or emotional intelligence remains unquestioned at the latest since then. Contemporary society is searching for alternative ability concepts that can explain success in academic, work, and private life. The present day zeitgeist no longer opposes cognitive reasoning to emotional or social skills, but rather values the contributions of both to a broad understanding of human resources (Matthews et al., 2002). Laypersons’ conceptions of human intelligence include other than traditional intelligence concepts, among them social competence (Sternberg, Conway, Ketron, & Bernstein, 1981; see also Kosmitzki & John, 1993). Most strikingly, personnel selection in all types of professions has already included the so-called soft skills into their selection strategies, at the same time hardly relying on established selection instruments to measure these skills.

The objective of the present work was to put social intelligence on a sound fundament in terms of a systematic integration of concepts, definitions, and theories. Moreover, the underlying methodological accounts and measurement approaches should be elaborated and applied to develop a test battery of social intelligence as a multidimensional cognitive performance construct. In the first Chapters (3-4), the central theories and concepts of intelligence research including academic and social intelligence as well as the related concepts of emotional and practical intelligence will be introduced. Chapter 5 will focus on the assessment of social intelligence including a systematic overview and description of past measurement approaches and a discussion of the pertinent failures and actual problems in the assessment of social and emotional abilities. Derived from these considerations, in Chapter 6, the aims of the present work, the test construction principles and procedures, and the research questions and hypotheses are formulated. Chapters 7 and 8 deal with the two main studies underlying the present work. These chapters include information about the samples, the material applied, and the results of the studies. Finally, Chapter 9 will discuss the utility of the test battery, and the scope, utility, and the validity of the social intelligence construct. In Chapter 10, some considerations about necessary research questions that add to the current results will point towards future perspectives.

3 Basic Theoretical Concepts, Terms and Definitions

In the encyclopedia, the term "intelligence" is defined as the abilities of the human intellect in the sense of an individual's potential and the consecutive dynamic meanings (Dorsch "Psychologisches Wörterbuch" Häcker & Stapf, 1994). Carroll (1993) defined intelligence as a generic term for cognitive ability constructs that are generally valid for the accomplishment of heterogeneous tasks, problems, and situations. Common to all broad theoretical models and definitions is the basic idea that academic intelligence is a cognitive performance construct. Construct definitions only include cognitive operations, and task requirements are independent from contextual or situational influences. An overview of definitions was given by Sternberg and Berg (1986) who presented the results of two symposia on intelligence which had taken place 65 years apart from each other in the years 1921 and 1986. In these symposia, the main protagonists of intelligence research were asked to define their notion of intelligence. Answers were manifold and ranged from intelligence is "what is valued by culture" to "speed of mental processing" with a maximum agreement of 57 % in 1986 (50 % in 1921) for the definition of "higher level components (abstract reasoning, representation, problem solving, decision making)" (p. 158). Common to most of the definitions was the idea of intelligence as a cognitive performance construct.

These symposia (Sternberg & Berg, 1986) already revealed interests in the extension of existing intelligence concepts. In 1921, 50 % and still in 1986, 25 % of the voters mentioned the "investigation of abilities other than cognitive" (p. 161) as one of the crucial next steps in intelligence research. Another 36 % in 1921 (21 % in 1986) asked for the investigation of *intelligence in specific domains (music, arts, chess)*. Furthermore, the *real-life manifestations of intelligence* were of interest for 14 % of the voters in 1921 and for 21 % in 1986. Furthermore, the extension of intelligence has been the topic of large amounts of diversely oriented efforts, among them the convention of reputable experts in intelligence research for a symposium on *The Enhancement of Intelligence* in the year 2001 (3rd International Spearman Seminar, Sydney, Australia; Kyllonen, Roberts, & Stankov, 2007). The contributions at this symposium reflected the diversity of extensions to intelligence concepts ranging from reductionist approaches to a "trend of diversification" (Süß, 2001, p. 109). A prototypical example of a reductionist approach is the mental speed paradigm (Neubauer, 1995; Vernon, 1993), which assumes that performance in intellectual tasks can be explained by and reduced to the general speed of information processing. At the other end of the spectrum, the introduction of a new ability construct represents an attempt to diversify the field of human intelligence.

In 1920, social intelligence was intended to extend traditional intelligence concepts. However, under the label of social intelligence, research was not as programmatic as, and was much more diverse than, academic intelligence research. Social skills were assessed as social intelligence by the use of self-report inventories (Marlowe, 1986; Riggio, 1986). Social behavior was judged by trained observers and should also operationalize social intelligence (Ford & Tisak, 1983, Frederiksen, Carlson, & Ward, 1984). Thus, compared to academic intelligence, the operationalizations also contained additional or distinct criteria than just cognitive requirements. The diversity of approaches resulted in limited progress of establishing social intelligence as a meaningful and unitary factor of human abilities. Moreover, the unsystematic use of definitions and measurement concepts resulted in legitimate skepticism of some authors (Ford, 1994) about whether to specify social intelligence as a *performance* or *ability* construct.

Intelligence, Competence, and "Performanz"

According to Weber and Westmeyer (2001), psychometric intelligence is considered a psychologically defined concept based on Wiggins' (1973) distinction between psychologically and socially defined concepts. The former are defined by psychological research and prototypically applied as predictors in the context of psychological assessment (e.g., academic intelligence). Underlying the idea of socially defined concepts is a social-constructivist perspective. In this respect, lay-psychologists in applied settings are responsible for defining and specifying a concept. Prototypical examples are external criteria in applied settings such as success in a job which is rated by a supervisor. Weber and Westmeyer criticized that these former external criteria are now introduced as new ability constructs; emotional (Salovey & Mayer, 1990), practical (Wagner & Sternberg, 1985), or successful intelligence (Sternberg, 1997) are, according to the authors, the ideal examples.

The expressions *intelligence* and *competence* were often applied as synonyms in social intelligence research. Süß, Weis, and Seidel (2005b) identified important distinctions between the two concepts. According to the authors, *competence* is specific to different situations and contexts (i.e., in certain applied settings) and more subject to modification and learning than *intelligence*. Intelligence is comparatively stable over time and seen as hereditary to a substantial extent (Grigorenko, 2000). Thus, following Weber and Westmeyer (2001), social *competence* can be classified as a socially constructed concept, as it comprises all person-related preconditions to show successful behavior in varying types of applied settings. Definitions of social competence vary substantially according to the spectrum of covered human attributes from just one (e.g., management of conflict, communication skills) to a complex interaction of various variables (Süß et al., 2005b). Intelligence often is one necessary part of competence concepts.

"Performanz", an expression especially known from German research literature, indicates the finally expressed *behavior* (= the result) in contrast to the person-related *preconditions* that "only" enable behavior (= the *potential*). Whether a person is capable of showing successful or effective behavior is not a direct function of this person's potential (i.e., competence and intelligence) and additionally, it is dependent from certain personality traits (e.g., shyness, altruism, etc.), from moods and current psychological states (e.g., fear, exhaustion), and from context variables (e.g., group values) (Süß et al., 2005b). The distinction between competence and "Performanz" is not only theoretical, it is also apparent when distinguishing between potential- and results-oriented approaches used to assess social competence. Contrary to the potential-oriented approaches, results-oriented approaches conceive social competence as effective behavior (= the outcome) where effectiveness is determined through the specific properties of the situation.

Abilities and Skills

Intelligence constructs usually consist of several distinguishable ability factors, for example, reasoning or verbal abilities. *Competence* constructs also contain cognitive and behavioral skills. According to Süß et al. (2005b, see also Scherer, 2007), skills are concrete actions or applications of cognitive operations on concretely defined problems (e.g., driving with a stick shift or applying an algorithm on some new data). Skills are acquired in a process of several steps and are finally characterized by an automated series of action (Ackerman, 1987). Contrarily, abilities represent more general, dispositional capacities. They are "either genetically endowed or acquired over a long period of socialization" (Scherer, 2007; p. 103).

Criteria to Judge Performance

Literature in the context of academic intelligence provides numerous accounts on the scoring of performance (Guttman & Levy, 1991; Nevo, 1993). Answers are scored according to well defined and reproducible rules so that the verity of a defined correct answer is not questioned (Roberts, Zeidner, & Matthews, 2001). A detailed description of general scoring principles and the problem of objectivity in the assessment of new ability constructs are provided in Chapter 5.2.4. However, things are not so clear when scoring other types of tests, such as those that are behavioral in nature. Scoring these types of tests are often more complicated since various criteria can or must be applied. These criteria can be classified into those that rely on the effectiveness of the achievement of a given goal, and those that judge behavior according to the compliance with valid social norms (Süß et al., 2005b). With respect to the latter, behavior is judged as socially competent when both the applied means and the achieved outcome are accepted by the respective social group. This implies context dependency of the scoring criterion. Consequently, assessing social competence according to the acceptance criterion requires highly specialized measures since every specific social group, setting, or situation, has its own standards.

Integrating the Concepts: A Model of Socially Competent Behavior (Süß et al., 2005b)

In summary, social competence can be classified into potential-oriented and results-oriented concepts. Table 3.1 contrasts the three concepts with the help of prototypical examples of German and Anglo-American research literature (extracted from Süß et al., 2005b).

Table 3.1

Models of Social Competence (Extracted from Süß et al., 2005b)

	Greif (1987)	**Kanning (2002)**	**Schneider, Ackerman, and Kanfer (1996)**
Definition	Effective realization of plans and aims in social interaction.	Universe of a person’s knowledge, abilities, and skills that promote socially competent behavior. The latter is defined as effective behavior specific to the context in accordance with the social group.	Effective social behavior and its cognitive, affective, and behavior-related preconditions.
Attributes / Dimensions	Social perception Interpretation of social cues	Social perception Behavioral control Assertiveness Social orientation Communication skills	Social intelligence Social skills Interpersonal personality traits Social self-regulation
Potential vs. Result	Result-oriented	Mixed model	Potential-oriented
Context included	Yes	Yes	No
Performance criterion	Efficiency	Efficiency and social acceptance	Efficiency

In a pure results-oriented approach, Greif (1987) applied the idea of a control loop on socially competent behavior. In this feedback loop, a person acts to reach a certain aim and compares the outcome with the desired state until the aim is achieved. In contrast, Kanning (2002) incorporated both a results-oriented approach relying on a control loop, and a potential-oriented approach with person-related attributes that contribute to socially competent behavior (e.g., social perception, assertiveness, etc.). Schneider et al. (1996) focused solely on person-related attributes in a potential-oriented approach including both, ability and personality variables.

In order to complete the conceptual framework for the present work, the integrative model of socially competent behavior by Süß et al. (2005b) will be described (see Figure 3.1).

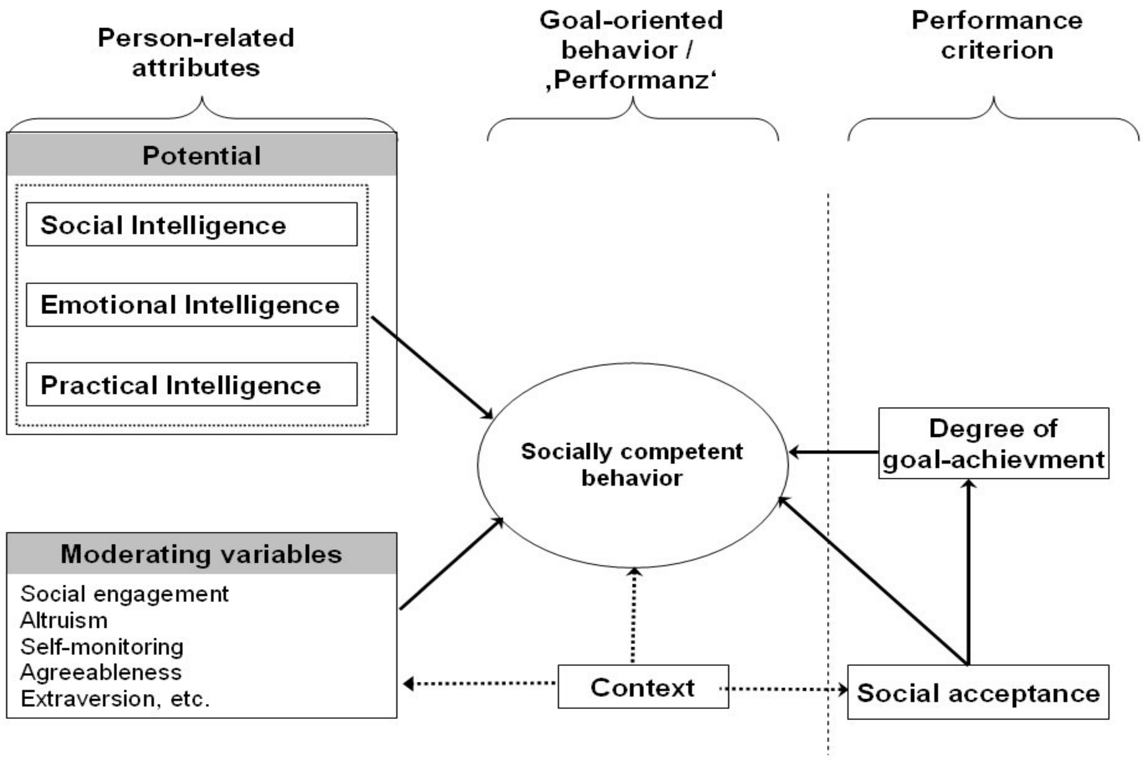


Figure 3.1

Integrative Model of Socially Competent Behavior (Adapted from Süß et al., 2005b)

The cognitive ability constructs of social, emotional, and practical intelligence belong to the person-related preconditions. Social competence additionally consists of social and emotional skills that are not depicted in the diagram. Furthermore, variables such as interpersonally relevant personality traits or interests (e.g., agreeableness, altruism, etc.) are supposed to moderate the “Performanz” of socially competent behavior. Altogether, Süß et al. (2005b) speak of the person-related attributes, or potential, of a person.

This framework integrates the preceding considerations in one diagram (see Figure 3.1). At the same time, it invalidates Ford’s (1994) critique that social intelligence cannot be specified as a pure ability construct (see also Asendorpf, 2002). According to Ford (1994) and Asendorpf (2002), individual differences in socially or emotionally intelligent performances cannot be specified without considering situational demands, social values, and personal aims. The present model of Süß (2005b) and colleagues clarifies the necessity to differentiate between the fundamental cognitive ability structure (the potential) as the preconditions that allow or influence the final social behavior, and the behavior itself (result). The moderating variables are expected to influence behavior directly and indirectly via an interaction with context-specific variables. For example, a certain context could reinforce the effect of a person’s interest and thus, may increase the probability of a concrete action. Cognitive ability constructs are, according to Carroll (1993), generally valid for the accomplishment of heterogeneous tasks, problems, and situations and thus, per definition, not context-dependent. Therefore, cognitive ability constructs should incorporate the ability to accomplish heterogeneous situational demands. In contrast, social behavior is influenced directly by the context and is supposedly always directed towards a social goal. Whether social behavior is judged as competent depends on whether or to what extent the goal is achieved and if the behavior and the goal comply with social norms.

The focus of the present work is the cognitive ability constructs with the implications that it is conceived as generally valid across situations, not related to specific situative demands, and without behavioral components in theoretical or behavior definitions.

4 Human Abilities – Traditional and New Intelligence Constructs

The following chapters deal with the theories and definitions of traditional and new intelligence constructs. First, some of the most influential theories of academic intelligence are presented with a focus on the type of the theoretical model, their relationship to each other, and the relevance for the social intelligence construct. Against this background, a systematic list of requirements for an intelligence construct will be assembled. Afterwards, definitions and theories of social intelligence are presented. A special focus is placed on the cognitive components of social ability concepts, resulting in a cognitive performance model of social intelligence. Lastly, new ability constructs such as emotional and practical intelligence will be presented, which are purportedly related to social intelligence. The type of overlap will be elaborated based on the presented theoretical foundations.

4.1 Theories of Academic Intelligence

In 1904, Spearman detected covariations between different types of sensory-discrimination tasks. Consecutively, he claimed to have identified a common general intelligence factor, the *g*-factor, which was supposed to be responsible for positive correlations among different kinds of mental tasks (Spearman, 1927). According to Spearman, the variance of a task can be explained, on the one hand, by the variance of the *g*-factor and, on the other hand, by the task-specific variance. This task-specific variance is not shared with any other task and hence cannot contribute to the covariance between two tasks. Consequently, the larger the relative proportion of *g*-related variance in two tasks, the larger the correlations should be among the tasks (i.e., the positive manifold). This finding was the origin for the so-called *g*-theory of which Jensen (1997) is the most prominent representative. Its basic assumption views a broad general intelligence factor as the only systematic source of common variance between mental tasks.

According to the *g*-theorists, these postulations have an important implication for new ability constructs. Unless a new construct shows substantially positive correlations with other measures of *g*, there is no justification to conceive it as a new intelligence construct (Austin & Saklofske, 2005; Gottfredson, 2003; Neubauer & Freudenthaler, 2005). Jensen's (1997) theory is subject to meaningful criticism from the intelligence literature (Bowman, Markham, & Roberts, 2002; Hedlund & Sternberg, 2000). Some critical points include that: *(a)* Many variables yield positive correlations with intelligence tests (e.g., openness for experiences, conscientiousness, etc.) and do not imply *g* as the source of common variance. *(b)* The *g*-factor represents the first principal component extracted in factor analysis, regardless of what intelligence test variables are entered into the analysis. Thus, the underlying functional commonalities of every *g*-factor depend on the selection of tests. *(c)* The amount of explained variance by the first principal component is often not higher than 50 % (Carroll, 1992). The extraction of a multiple-factor solution often results in larger variance explanation. *(d) g* was intended to represent the systematic variance common to all types of mental tasks. In practice, *g* is often measured by only one single variable (e.g., Raven's Progressive Matrices Test) (Bowman et al., 2002). In summary, *g* in the absence of further lower-order ability factors could also be interpreted as a mathematical artifact; any conclusions based on the primacy of *g* should be derived with care.

In contradiction to *g* as one single source of common variance, Thurstone (1931) detected seven primary mental abilities that could explain correlations between mental tasks. In a factor analytic approach, he identified verbal comprehension, word fluency, number facility, spatial visualization, associative memory, perceptual speed, and reasoning as separate ability factors. These factors only emerged in the absence of a common general factor. Accordingly, Thurstone formulated the first structural model of human intelligence that differentiated between several more or less broad ability domains. Albeit apparently contradictory, combining the two approaches unveils a hierarchical theory of human intelligence which postulated a higher-order general factor and lower-order primary ability factors (Brody, 2000). To establish a hierarchical model, the respective lower-order abilities (in this case Thurstone's primary abilities) need to be intercorrelated in order to allow a superordinate factor structure on the next higher level in the hierarchy. The assignment of the respective lower-order abilities to higher-order factors is supposed to be derived from theory-based assumptions about the task requirements. In subsequent research, several structural and hierarchical models of human intelligence were introduced. All of them were concerned with decomposing variance in mental tasks into various sources, that is, *g*-related variance or variance related to some broad or more specific ability factors. Thurstone (1931) and Guilford (1967) formulated structural models of intelligence whereas Cattell (1963, 1971), Carroll (1993), and Jäger (1982) have established hierarchical models.

Guilford's Structure of Intellect Model (SOI; 1967)

Guilford (1967) postulated a structural theory of intelligence. His Structure of Intellect Model (SOI) represents a faceted model of intelligence that cross-classifies the three facets of *operations*, *products*, and *contents* (for another faceted model of intelligence, see the last section in the present Chapter). The facet o*perations* describes the cognitive requirements people need to accomplish a task and it contains five elements: cognition, memory, divergent production, convergent production, and evaluation. The *content* facet includes four elements and refers to the properties of the task material: figural, symbolic, semantic, and behavioral. Finally, the *product* facet comprises six elements, each describing a type of outcome associated with a mental task: units, classes, relations, systems, transformations, and implications. Consequently, the cross-classification of the facets resulted in 120 postulated ability factors that, according to Guilford (1967), described distinct human intellectual abilities. In a first adaptation, Guilford (1977) divided the figural contents into auditory and visual contents which resulted in additional 30 factors (see Figure 4.1 for the complete SOI).

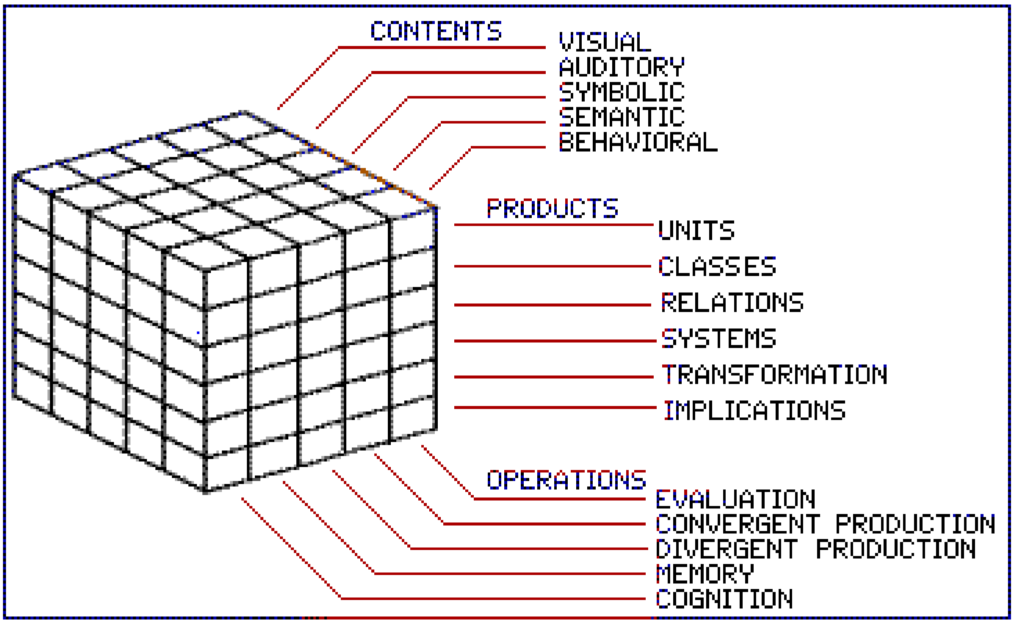


Figure 4.1

Structure of Intellect Model (Guilford, 1967, 1981)

Guilford later modified the basic assumption of orthogonal factors (Guilford, 1981) and allowed intercorrelations between the ability factors. Thus, the modified SOI represents a hierarchical model with 150 first-order factors, 85 second-order factors, and 16 third-order factors. However, empirical data could never fully confirm the postulated factor structure (Brody, 2000).

Notably relevant for the social intelligence construct, the SOI also included a subset of abilities that were identified as representing social intellectual abilities, namely, all 30 cross-classifications with the behavioral contents. Behavioral contents were defined as “essentially nonverbal information, involved in human interactions, where awareness of attention, perceptions, thoughts, desires, feelings, moods, emotions, intentions, and actions of other persons and of ourselves is important” (Guilford, 1967, p. 77). Guilford described behavioral information as mostly conveyed via visual or auditory cues. For testing purposes, he relied mainly on pictorial material. This seemed to cause a conceptual overlap with the figural and also the later auditory content domain. He acknowledged a possible role of the figural ability factors for the accomplishment of tasks based on the behavioral content presented in pictures. But at the same time, he stressed that the content (“the substance”; p. 221) of the behavioral information is essential and not the form of stimulus presentation. Thus, Guilford’s (1967; 1981) model established social intelligence as independent from any content-related differentiations. Some of the 30 ability factors of social intelligence were operationalized by sets of specifically developed tasks (Hendricks, Guilford, & Hoepfner, 1969; O’Sullivan & Guilford, 1966, 1976) and investigated by rather extensive empirical analysis (O’Sullivan, Guilford, & DeMille, 1965; Hendricks et al., 1969). The test batteries and the most important associated empirical results will be presented in a later section.

Cattell’s Theory of Fluid and Crystallized Intelligence (1963, 1971)

Cattell (1987) conceived the identification of the structure of intelligence as the first goal of ability research. He acknowledged the general value of Guilford’s approach in classifying intellectual abilities. However, he conceived the idea of unrelated ability factors as postulated in Guilford’s early model as “entirely wrong” (p. 37). Cattell proposed a hierarchical theory of intelligence with two correlated general factors at the apex of the hierarchy, namely, fluid and crystallized intelligence (g_f and g_c respectively) (Cattell, 1963, 1971). Fluid intelligence reflects the ability to deal with abstract information in tasks like series, classifications, analogies, etc. Importantly, fluid abilities explicitly exclude knowledge requirements (i.e., task material must be equally accessible for any tested group so that the opportunity of knowledge acquisition is the same for everyone). Fluid abilities are supposed to be hereditary and to decline over the lifespan. Crystallized intelligence shows loadings of, for example, verbal and numerical skills, and reflects knowledge acquired over the lifespan with no age-related decline. Cattell (1987) reported correlations between g_f and g_c of .40 to .50. He reasoned in accordance with his Investment Theory that the acquisition of crystallized abilities depends on the investment of fluid abilities. Cattell (1987) later extended the theory by adding three broad factors besides g_f and g_c: retrieval capacity (g_r), visualization (g_v), and cognitive speed (g_s). On a lower level in the hierarchy, over 40 first-order factors were postulated. Contrary to other theories of human intelligence, Cattell's theory offered considerations and testable hypotheses about the intellectual development across the lifespan. Empirical studies supported the hypothesized developmental trajectories of the different abilities (Schaie, 1996). In German literature, the Intelligence Structure Test (IST-2000 R; Amthauer, Brocke, Liepmann, & Beauducel, 2001) is based on the g_f-g_c-theory. Comprehensive empirical support for the entire theory cannot be found in the literature.

Carroll's Three-Stratum-Theory (1993)

Carroll's Three-Stratum-Theory (1993) conceptually relies on Cattell's (1987) g_f-g_c-theory and also represents a hierarchical theory. The theory is based on an exhaustive reanalysis of correlative datasets from psychometric intelligence research. The theory shows three levels of generality (called strata): Stratum III represents the highest level of the hierarchy and consists of a general intelligence factor. On Stratum II, eight different ability factors are located that differ in terms of their loadings on the general intelligence factor. In the order of high to low loadings on the *g*-factor, the ability factors are: *fluid intelligence*, *crystallized intelligence*, *general memory and learning*, *broad visual perception*, *broad auditory perception*, *broad retrieval ability*, *broad cognitive speediness*, and *processing speed*. Comparable to the g_f-g_c-theory, Stratum I includes numerous specific abilities (see Figure 4.3 for a display of the three strata). In an integrative approach, the g_f-g_c-theory (Cattell, 1973, 1971) and the Three-Stratum-Theory were just recently combined in the Cattell-Horn-Carroll-Theory (CHC) (McGrew & Evans, 2004). The Woodcock-Johnson Psychoeducational Battery (WJ-III; Woodcock, McGrew, & Mather, 2000) was constructed to follow the propositions of the g_f-g_c-theory combined with some Stratum-II-factors of Carroll's theory.

Berlin Intelligence Structure Model (Jäger 1982, 1984)

The Berlin Intelligence Structure Model (BIS-Model; Jäger 1982, 1984) also represents an integrative hierarchical model that differs from the aforementioned hierarchical approaches in some pivotal aspects. The BIS-Model was derived from several large empirical studies that applied a nearly exhaustive collection of existing intelligence tasks. Tasks that were excluded were those that were redundant to tasks already included in terms of their cognitive determinants. Furthermore, marker variables for competing intelligence factors were kept in the analysis (Süß & Beauducel, 2005). Contrary to Cattell's (1963, 1971) and Carroll's (1993) theory, the idea of a faceted model of intelligence is applied in this model. The resulting model proposes a general factor at the top level of the hierarchy and seven second-order factors. These belong to two facets: an operational facet that consists of four cognitive operations (i.e., *reasoning*, *creativity*, *memory*, and *speed*) and a content-facet consisting of three content-related ability factors (i.e., *numerical*, *figural*, and *verbal abilities*). The cross-classification of the two facets results in 12 cells that represent a third, more specific level in the hierarchy. Figure 4.2 displays the BIS-Model showing all three levels of generality. Contrary to Guilford's (1967) faceted model of intelligence, the cells do not represent separate ability factors but serve as a classification system for tasks (Süß & Beauduce

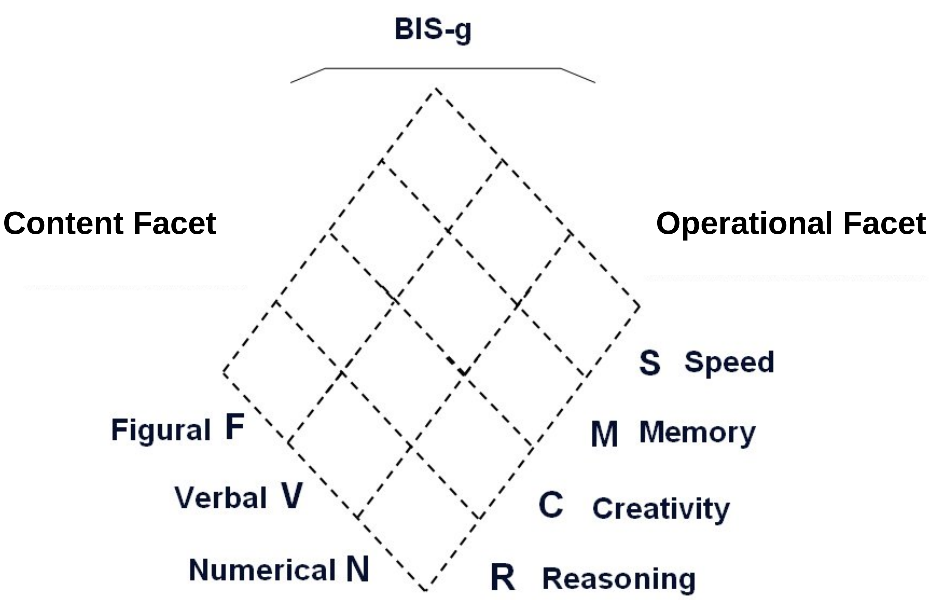


Figure 4.2
Berlin Intelligence Structure Model (Jäger, 1982, 1984)

According to Jäger's conceptualization, knowledge requirements (i.e., as specified in g_c) are represented in the content-facet. *Basic knowledge* about, for example, vocabulary or mathematical rules is necessary to accomplish verbal or numerical tasks. This distinguishes the faceted BIS-Model from any non-facetted model such as Cattell's theory. In the BIS-Model, a content-related ability facet balances any additional requirements related to the task material and any associated basic knowledge demands.

The seven second-order ability factors are defined as follows:

On the operational facet:

- *Speed (BIS-S)*: the ability to accomplish simple tasks quickly and accurately
- *Memory (BIS-M)*: the ability to recall lists and configurations of items
- *Creativity (BIS-C)*: the ability to fluently produce many different ideas
- *Reasoning (BIS-R)*: the ability to elaborate complex information including inductive and deductive reasoning, construction, and planning

On the content facet:

- *Verbal (BIS-V)*: the ability to deal with verbal material using the four different cognitive functions
- *Figural-Spatial (BIS-F)*: the ability to deal with figural-spatial material using the four different cognitive functions
- *Numeric (BIS-N)*: the ability to deal with numeric material using the four different cognitive functions

Thus, the benefit of a faceted model in comparison with a structural model concerns the possibility to decompose task variance into the operational and content-related abilities of the two facets. This allows an unbiased analysis of the covariance structures between tasks. By aggregating tasks across an ability domain of one facet, variance due to abilities of the other facet is controlled for. If this is not accomplished, for example, *fluid abilities*, *reading comprehension*, and *spatial scanning* in the Three-Stratum-Model as First-Stratum factors are assigned to different Second-Stratum factors although it can be assumed that fluid abilities rely on reading comprehension when verbal material must be dealt with and on spatial scanning when figural spatial material must be dealt with.

The BIS-Model represents the theoretical foundation for the Berlin Intelligence Structure Test (BIS-Test; Jäger et al., 1997). The internal structure of the BIS-Model was replicated in various studies (Beauducel & Kersting, 2002; Brunner & Süß, 2005; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Figure 4.3 illustrates a classification of the BIS-ability domains into the Three-Stratum-Model. Except for Broad Auditory Perception, for every Stratum II ability factor, there is an equivalent ability factor in the BIS-Model. This classification, however, suffers from a mismatch between the faceted BIS-Model and the non-faceted Carroll model. *Stratum I* primary abilities classified to operative broad ability factors and *Stratum II* comprise unbalanced verbal, numerical, or figural material-related components (e.g., *word fluency* as a primary ability subsumed under *broad retrieval ability*, *numerical facility* subsumed under *broad cognitive speediness*). Moreover, the BIS-Model lacks a content-factor related to auditory abilities which is represented in the Three-Stratum-Model. Auditory abilities are seen as a relevant complement to content-related abilities (Stankov, 1994; Stankov & Horn, 1980). Seidel (2007) applied a first attempt to integrate an auditory ability factor into the BIS-Model and developed several new performance tests based on the battery of Stankov and Horn (1980).

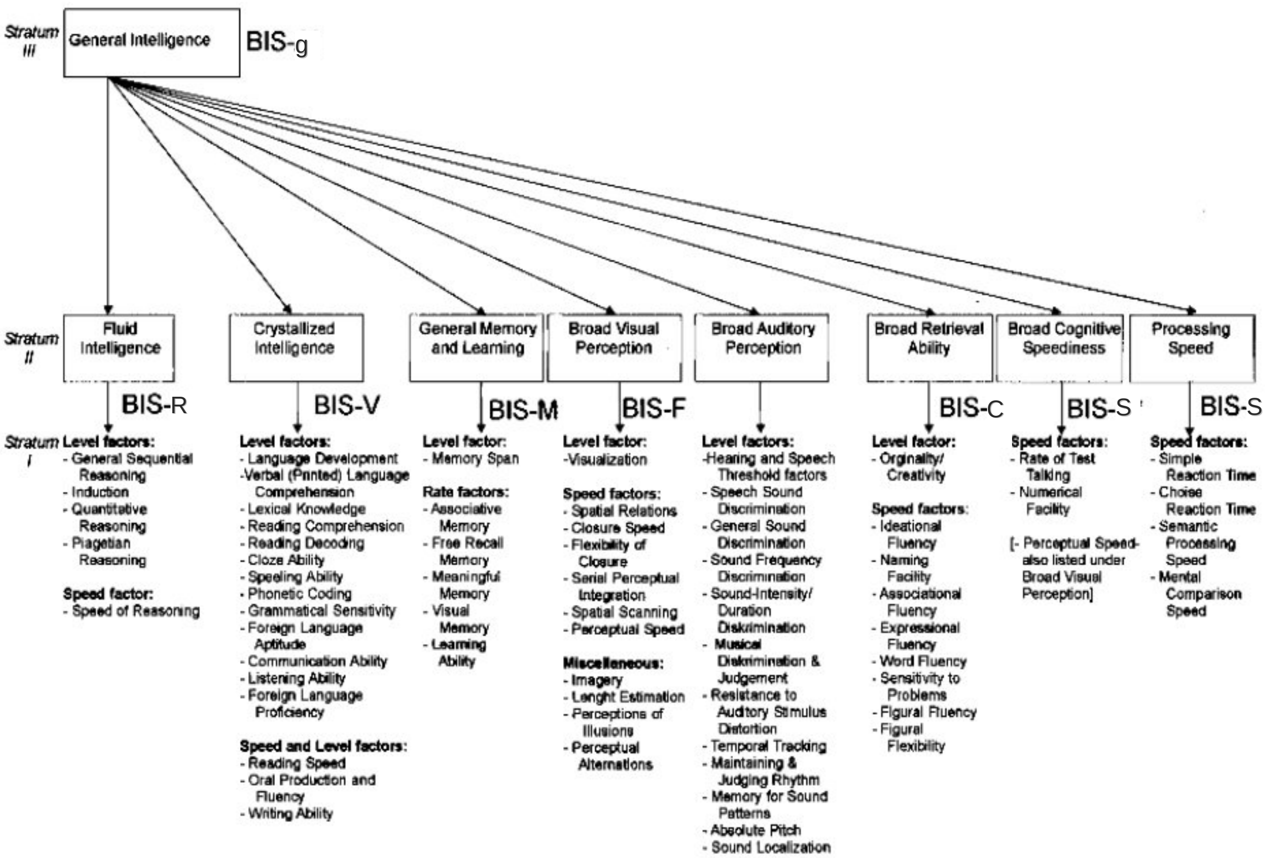


Figure 4.3

Three-Stratum-Model and the Ability Factors of the BIS-Model (Carroll, 1993, p. 626; Süß, Seidel, & Weis, 2005a)

Note. R = Reasoning, V = Verbal abilities, M = Memory, F = Figural, C = Creativity, S = Speed

4.2 Requirements for a New Cognitive Ability Construct

Social and also emotional intelligence represent new candidates as meaningful intelligence constructs. However, the conceptual and methodological problems surrounding the introduction of a new intelligence construct are manifold (Landy, 2006; Matthews et al., 2005; Schaie, 2001; Weber & Westmeyer, 2001). Many prominent intelligence researchers established requirements for the introduction and foundation of a new intelligence construct. The following section will be concerned with integrating these various positions.

a) Theory formation

These requirements refer to the nature and the development of underlying theoretical foundations. Matthews et al. (2005) required a-priori theoretical considerations about the "localization of the intended construct within the sphere of individual differences" (p. 80) and, in this respect, coherence in theory (i.e., theory has to provide postulations about the nature of the coherence between the construct and its manifestation in behavior). O'Sullivan (1983) called for a clear and nonredundant terminology for construct specification. Süß (2001) demanded a construct definition based on empirical results. In summary, research on a new intelligence construct needs to clarify the concepts, labels, and scope to establish the conceptual fundament.

b) Construct specification:

The respective construct needs to be specified in terms of the underlying operational performance determinants. This also implies the definition and classification of possible subconstructs in the context of hierarchical theories and hypotheses about the pattern of relations among them. Carroll (1993) claimed that intelligence must consist of cognitive determinants only. Süß (2001, 2006) required that the construct consists of highly general abilities that are relatively stable over time. Consequently, construct definitions should cover a broad and heterogeneous range of ability and content domains. Moreover, Süß (2001, 2006) demanded a minimum amount of knowledge requirements as a necessary prerequisite for a generally valid intelligence construct.

c) Operationalizations

Requirements concerned with the assessment of the new construct refer to both the formal characteristics of the measurement approach (i.e., type of data and scoring), and the psychometric qualities of the respective diagnostic instruments. With respect to the psychometric qualities, Matthews et al. (2005) and Weber and Westmeyer (2001) stressed the necessity of psychometrically sound operationalizations (i.e., item properties and reliability). With respect to the former point, O'Sullivan (1983) addressed the necessity of congruence between the construct specification and the tasks chosen to measure it. Süß (2001, 2006) and Weber and Westmeyer (2001) demand the application of true performance tests in the sense of T-data according to Cattell (1965). Furthermore, objective scoring rules have to be available in order to judge the performance in tasks.

d) Validation

For proving the validity of the new intelligence construct, the validation strategy must be selected carefully. Both convergent and discriminant construct validity has to be proven by data (O'Sullivan, 1983; Weber & Westmeyer, 2001). Schaie (2001) requested the selection of adequate measurement procedures for the assessment of the validation instruments (Weis & Süß, 2005). For example, academic intelligence should be measured by the assessment of a multidimensional test battery in order to investigate the differential pattern of correlation with fluid or crystallized intelligence, speed, memory, verbal, figural-spatial, or numeric abilities. With respect to the criterion validity, studies have to provide evidence for the incremental predictive validity for heterogeneous external criteria over and above traditional predictor variables to prove the relevance and the practical meaning (i.e., academic intelligence, personality traits) (Süß, 2001, 2006).

e) Other requirements

Austin and Saklofske (2005) additionally required evidence for the genetic determination of an individual's performance (i.e., evidence for the heritability of the new intelligence construct) and evidence for biological manifestations specified by basic neurological measures (e.g., speed of information processing). Whether the genetical determination must be proven empirically, is not specified. However, given the aforementioned criteria, genetical determination can be assumed to some extent. Schaie (2001) specifically addressed research on emotional intelligence and demanded positive correlations with age (i.e., experience) that could be inferred from construct definitions.

One point is especially controversially discussed. Conforming with Jensen's (1997) idea of the primacy of a general intelligence factor and the proclaimed positive manifold, Austin and Saklofske (2005), Gottfredson (2003), Guttman and Levy (1991), Mayer et al., (2000), and Neubauer and Freudenthaler (2005) all consider positive correlations between traditional and new intelligence tasks as essential for the establishment of a new intelligence construct. According to Henry, Sternberg, and Grigorenko (2005), multiple types of intelligence do not have to fit under one single general factor. Also Bowman et al. (2002) identified meaningful criticisms of this requirement. The present work does not see evidence for the positive manifold as a necessary condition for the introduction of a new intelligence construct. First of all, the covariance pattern highly depends on the presumed level in the hierarchy in both constructs. Moreover, it is considered more important that the aforementioned commonly accepted criteria are accomplished.

4.3 Theories and Definitions of Social Intelligence

4.3.1 Definitions of Social Intelligence

The only comprehensive theory based account of social intelligence is Guilford's (1967) Structure of Intellect Model. In the SOI, the behavioral content domain reflects social intelligence. It is located aside figural, symbolic, and semantic task contents, which implies that social and general intelligence, according to Guilford, only differ in the contents of the tasks but share the same cross-classifications with operations and products. Thus, social intelligence includes cognition, convergent and divergent production, memory, and evaluation of behavioral contents. These contents mostly consist of nonverbal information about social interactions that allow conclusions about thoughts, desires, feelings, moods, emotions, intentions, and actions of other persons and of ourselves. Guilford and his colleagues (Hendricks et al., 1969; O'Sullivan et al., 1965) focused on the operational domains of cognition and divergent production to develop tests of social intelligence. O'Sullivan and Guilford's efforts resulted in two test publications, the Six Factor Test (O'Sullivan & Guilford, 1966) and the Four Factor Test (O'Sullivan & Guilford, 1976) of Social Intelligence. Other definitions of social intelligence reflect Thorndike's (1920) early distinction between cognitive and behavioral requirements (i.e., "understanding people" vs. "act wisely in human relations") (p. 228). Table 4.1 lists definitions extracted from the literature. The list contains both definitions from theoretical accounts and operationalizations. They are classified into cognitive and behavioral components. Additionally, the cognitive components are subdivided into different operational requirements (i.e., reasoning, memory, perception, creativity, and knowledge requirements).

Table 4.1

Definitions of Social Intelligence Extracted from the Literature

<table>
<tr><th>Cognitive Requirements</th><th>Cognitive Components</th><th>Behavioral Components</th></tr>
<tr><td>Reasoning</td><td>Insight into the moods or personality traits of strangers (Vernon, 1933)
Judge correctly the feelings, moods, and motivation of individuals (Wedeck, 1947)
Ability to judge people with respect to feelings, motives, thoughts, intentions, attitudes, etc. (O'Sullivan et al., 1965)
Understand the feelings, thoughts, and behaviors of persons, including oneself (Marlowe, 1986)
Judgment in social situations (Moss et al., 1955)
Recognition of the mental states behind words and from facial expressions (Moss et al., 1955)
Role-taking ability (Feffer, 1959)
The ability to interpret social cues (O'Sullivan & Guilford, 1966)
The ability to predict what will happen (O'Sullivan & Guilford, 1966)
The ability to identify the internal mental states (O'Sullivan & Guilford, 1966)
Decoding of social cues (Barnes & Sternberg, 1989; Buck, 1976; Sundberg, 1966)
Ability to comprehend observed behaviors in the social context in which they occur (Wong, Day, Maxwell, & Meara, 1995)</td><td rowspan="5">Get along with others and ease in society (Vernon, 1933)
Ability to get along with others (Moss & Hunt, 1927)
The ability to deal with people and the applications of means to manipulate the responses of others (Orlik, 1978)
Act appropriately upon an understanding of the feelings, thoughts, and behaviors of persons, including oneself (Marlowe, 1986)
The ability to manipulate the responses of others (Weinstein, 1969)
Attainment of relevant social goals (Ford, 1982)
Ability to speak effectively, to be appropriately responsive to the interviewers questions, to display appropriate nonverbal behaviors (Ford & Tisak, 1983)
Effectiveness in heterosexual interaction (Wong et al., 1995)
Social problem solving (Cantor & Harlowe, 1994</td></tr>
<tr><td>Memory</td><td>Memory for names and faces (Moss et al., 1955; Sternberg et al., 1981)</td></tr>
<tr><td>Perception</td><td>Sensitivity for other people's behavior (Orlik, 1978)
The ability to perceive the present mood of other people (Orlik, 1978)</td></tr>
<tr><td>Creativity (Fluency)</td><td>The ability to create recognizable categories of behavioral acts (Hendricks et al., 1969)
The ability to imagine many possible outcomes of a setting (Hendricks et al., 1969)</td></tr>
<tr><td>Knowledge</td><td>Knowledge of social matters (Vernon, 1933)
The capacity to know oneself and to know others (Gardner, 1983)
Individuals fund of knowledge about the social world (Cantor & Kihlstrom, 1987)
Social problem solving (Cantor & Harlowe, 1994)
Knowledge of rules of social interaction (Orlik, 1978)
Knowing the rules of etiquette (Wong et al., 1995)</td></tr>
</table>

Reasoning requirements obviously represent the broadest domain, and are labeled social understanding from now on. Within this domain, the cognitive operations of *understanding*, *interpreting*, *judging*, *having insight*, *predicting*, and *comprehending* all address comparable or identical cognitive operations. In contrast, *recognizing*, *decoding*, and *identifying* seem to address different cognitive operations that require less information processing but rather equal initial (perceptual) functions. The definitions of person perception in social psychological research equal the definitions for social understanding in the present work, and are applied synonymously. In social psychological research, the concepts of person perception or judgmental accuracy received substantial attention and now belong to the most established concepts in social (cognitive) psychology. Research on social or interpersonal perception (Bronfenbrenner, Harding, & Gallwey, 1958; Cline, 1964), social cognition, or interpersonal processes (Weinstein, 1969) offers important and fruitful definitions, concepts, and measurement principles for the study of social intelligence (see Chapters 4.3.3.1 and 4.3.3.3 for a more detailed look on some interesting results from experimental studies).

Memory, perception, creativity, and knowledge requirements do not cover such heterogeneous functions. All five cognitive operations will be addressed in more detail in the forthcoming passages. In contrast, the definition of social intelligence as social problem solving to attain social goals (Cantor & Harlowe, 1994; Cantor & Kihlstrom, 1987) will be excluded. Social problem solving represents a special category with, on the one hand, less explicit cognitive requirements, and on the other hand, additional behavioral or knowledge components. According to Cantor and Harlowe (1994), problem solving strategies, schemas, and procedural rules for processing social information together represent the social intelligence repertoire reflected in knowledge structures. Schemas include concepts of oneself, others, and social situations. Because of this mixture of operative functions, social problem solving will be excluded from the performance model of social intelligence. Furthermore, operational definitions from approaches that rely on self-report measures of social intelligence are omitted. Some of these definitions are the same as those already derived from performance-based definitions (e.g., social and emotional sensitivity, Riggio, 1986; decoding ability, Zuckerman & Larrance, 1979). At the same time, other definitions reflect personality-like construct definitions such as empathy, social assertiveness, or self-efficacy (Marlowe, 1986).

4.3.2 A Cognitive Performance Model of Social Intelligence

Weis and Süß (2005; see also, Weis et al., 2006) proposed a performance model of social intelligence that incorporated the aforementioned structure of cognitive abilities. Besides the classification of cognitive operations, further taxonomic considerations of, for example, task contents will be addressed subsequently. The performance model was presented in both, an English and a German language book chapter (Weis & Süß, 2005; Weis et al., 2006, respectively). For reasons of simplicity, only the English language chapter will be referred to, whereby always the German language chapter would as well be valid.

Social Understanding

The core ability domain of social intelligence is social understanding. It includes cognitive operations subsumed under reasoning requirements (Weis & Süß, 2005). According to the authors, social understanding requires individuals to understand or interpret social stimuli against the background of the given social situation (e.g., understand correctly what a person wants to express via verbal or nonverbal means of communication). The stimuli can vary according to their complexity (e.g., from a simple facial expression to a sequence of interactions between persons) and should allow conclusions about a person's emotions, thoughts, intentions, motivations, or personality traits. The present definition excludes the more initial cognitive functions of recognizing or, in other words, perceiving social stimuli These are classified as social perceptual abilities.

Social Memory

So far, social memory was defined and operationalized as memory for names and faces (Kosmitzki & John, 1993; Moss et al., 1955). Guilford (1967) specified memory of behavioral contents as social memory. Weis and Süß (2005) defined social memory as the storing and recall of objectively given social information that can vary in complexity. For example, memory for names and faces is a narrow subset of social information (Probst, 1982), whereas the memory for a sequence of interactions represents a rather complex entity. The required social information has to be objectively present in the situation (i.e., in the task material respectively). However, presence itself does not insure that individuals direct their attention towards the relevant cues and thus, also perceive the cues. This problem carries two implications. First, the instructions for a task have to direct an individual's attention towards the relevant cues. Second, the type of information is supposed to influence the possibility for test takers to perceive the relevant cues. In this context, information included in static types of stimuli like written language and pictures suggest that all relevant cues can be perceived if the presentation time is long enough. Fluent information, for example, included in spoken language and videos only occur at one point in time and thus, are much harder to direct attention to. These last considerations reveal the importance of perceptual abilities discussed in the next section.

Social Perception

From a theoretical perspective, there is little doubt that social perception represents a relevant ability domain. Both social understanding and social memory require the prior perception of relevant stimuli (e.g., a person's smile needs to be perceived in order to make a conclusion about the person's mental state). In real life, this perception usually happens within a very short period of time or with restricted access to the relevant cues. Accordingly, Weis and Süß (2005) defined social perception as the ability to (quickly) perceive socially relevant information in more or less complex situations. Wong et al. (1995) operationalized social perception but could not separate this ability from social understanding abilities. In order to separate social perception from social understanding, two presuppositions should be achieved. First, the target stimuli have to be present in the situation. Thus, only overt behavior (or a predefined target stimuli) can be perceived (e.g., eye contact or a touch between two persons). Second, for the construction of adequate measures, Weis and Süß (2005) recommended the application of speed measures analogous to the concept of general perceptual (or mental) speed in theories of academic intelligence. By measuring social perception with reaction time scores, a higher-level information processing and a further elaboration of information is supposed to be eliminated or reduced from the score.

Social Creativity

Guilford (1967) introduced the *divergent production of behavioral contents* as one ability domain of social intelligence. Hendricks et al. (1969) constructed an unpublished test battery of this domain, where they defined social creativity as the ability to imagine possible outcomes of a setting or to create recognizable categories of behavioral acts. Recent empirical work (Jones & Day, 1997; Lee, Day, Meara, & Maxwell, 2002) operationalized social cognitive flexibility as the fluent production of possible interpretations of a social situation. The score represented the number and the diversity of given answers. In academic intelligence models, creativity is represented as retrieval abilities (e.g., broad retrieval abilities in Carroll's Three Stratum Model, or retrieval capacity g_r in Cattell's g_f-g_c-theory). In summary, Weis and Süß (2005) defined social creativity similarly like Lee et al. (2002) as the production of as many and as diverse solutions or explanations as possible for a social situation or problem.

Social Knowledge

In models of academic intelligence, knowledge plays diverse roles. Some models explicitly include knowledge as one ability domain of academic intelligence (Carroll, 1993; Cattell, 1971). Others do not account for a separate knowledge factor (Guilford, 1967; Jäger, 1982). At best, knowledge requirements should be eliminated as far as possible from ability domains that are different from knowledge itself (Cattell, 1987; Süß, 2001). Cattell (1987) acknowledged that performance in knowledge tasks always depends on an individual's learning experiences. However, in academic intelligence research, task contents are typically taught in standardized settings so that individuals' knowledge fundament presumably overlaps to a substantial extent. What is highly standardized in the academic knowledge domain, the learning environments for social knowledge vary substantially in terms of knowledge contents. Learning environments are the family, the peer group, school- or work-related groups, etc. The resultant knowledge contents depend on the cultural environment (Weber &Westmeyer, 2001). Every social entity (e.g., from a family consisting of father, mother, and child, to a culture or nation) has its own social standards so that possibly diverse social contents are taught as the "correct" knowledge. Thus, what is correct can only be judged with respect to the present social contexts and the social group. This implies that the construction of a social knowledge test requires a thorough definition of the social entity of which knowledge should be specified, a comprehensive classification of possible social situations within this entity, and subsequently, a rule according to which knowledge contents can be judged as right or wrong.

In the literature, social knowledge is defined as knowledge about the social world (i.e., social rules, social matters, etc.) (Cantor & Kihlstrom, 1987; Vernon, 1933). Also, operationalizations reflect knowledge as good etiquette in very specific settings (e.g., dinner-related knowledge; Lee, Wong, Day, Maxwell, & Thorpe, 2000; Wong et al., 1995). Altogether, social knowledge cannot be seen as a pure cognitive ability and does not fulfill the typical requirement of a general ability construct that is valid in heterogeneous situations. Furthermore, it seems impossible to develop and validate an adequate measurement instrument without applying a homogeneous sample with comparable experience in the queried knowledge domain. For these seasons, social knowledge is treated differently from the remaining cognitive ability domains in the upcoming considerations.

Integrating the operative ability domains

The present work distinguishes between the four cognitive ability domains of social intelligence as specified in the performance model of Weis and Süß (2005) (i.e., social understanding, memory, perception, and creativity) and acknowledges the relevance of social knowledge for social behavior within a framework of social competences. However, a crystallized social intelligence factor, as specified in academic intelligence models, is not subsumed under the cognitive performance construct in a narrow sense.

The performance model of Weis and Süß (2005) did not make a statement about the internal structure of social intelligence. They did, however, not exclude the possibility of correlated abilities and thus, of a higher-order general social intelligence factor. Furthermore, they assumed that all cognitive ability domains predict social behavior to a certain extent (see Figure 4.4). The present work extends and modifies the model of Weis and Süß and postulates positively related cognitive ability factors so that, on a higher-order level, a general social intelligence factor is postulated. Figure 4.4 and 4.5 display the two alternative performance models of social intelligence with *(a)* a structural model of social intelligence (Fig. 4.4; Weis & Süß, 2005) and *(b)* a hierarchical model of social intelligence (Fig. 4.5).

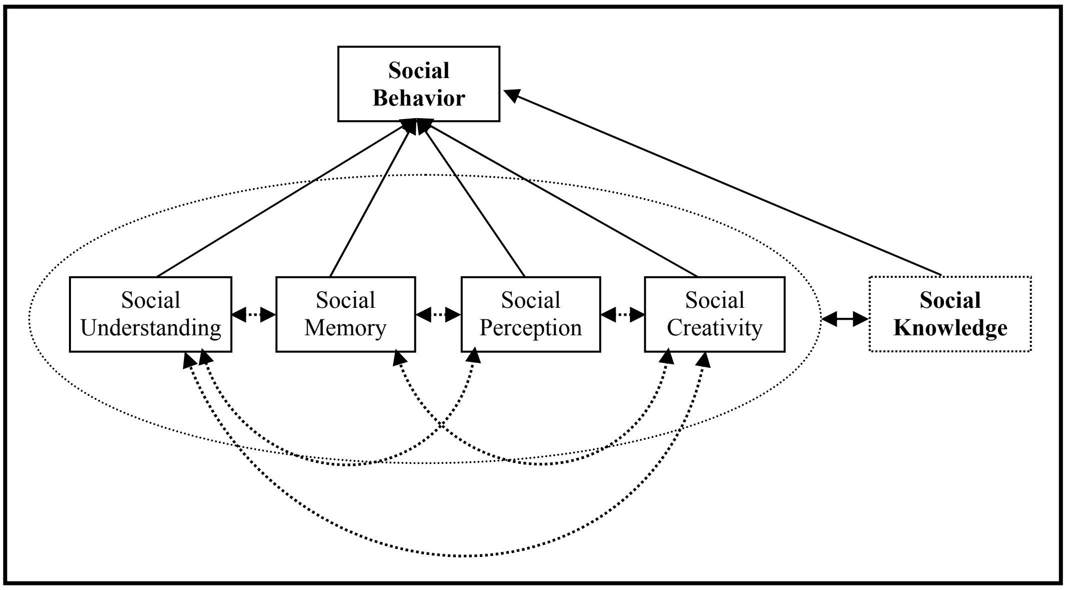


Figure 4.4

Performance Model of Weis and Süß (2005) Representing a Structural Model of Social Intelligence

In both models, the ability domains of social understanding, social memory, social perception, and social creativity constitute social (cognitive) intelligence in the narrow sense.

The structural model of Weis and Süß (2005) also classified social knowledge to the cognitive abilities subsumed under social intelligence. In the present hierarchical model, social knowledge is assigned a special role and it is assumed to be positively related to a putative general social intelligence factor. Both models claim to predict social behavior. For example, social behavior may be predicted by a general social intelligence factor (e.g., in typical private settings). At the same time, social behavior in a specific job setting could as well be specifically predicted only by social understanding (e.g., success in a counseling job). The model of prediction must be determined according to the principles of symmetry between predictor and criterion in order to optimize prediction (Wittmann, 1988; Wittmann & Süß, 1999). In this respect, symmetry refers to analogous hierarchical levels of analysis in terms of underlying hierarchical models of predictor and criterion variables.

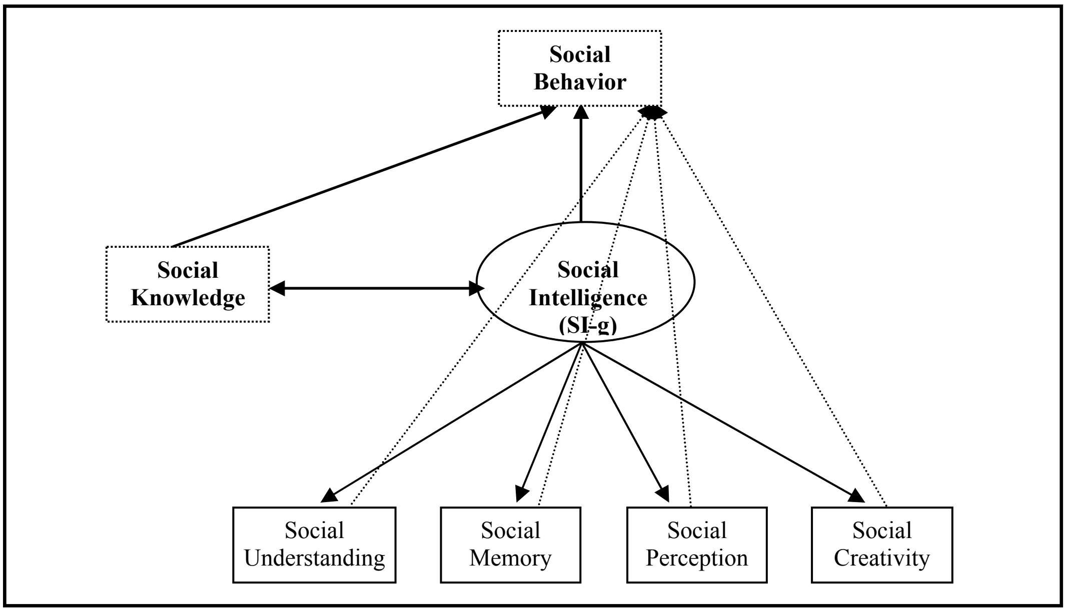


Figure 4.5
Modified Performance Model of Weis and Süß (2005) Representing a Hierarchical Model of Social Intelligence

The original performance model (Figure 4.4) was supported by empirical data in a multitrait-multimethod study of Weis and Süß (2007). Confirmatory factor analysis yielded the correlated factors of social understanding, social memory, and social knowledge and also supported a general social intelligence factor (see Chapter 5.4).

4.3.3 Taxonomic Foundations of Social Intelligence

For the establishment of an intelligence construct, Cattell (1987) demanded classificatory principles in the form of taxonomies. A taxonomy includes classificatory principles comparable to facets in the context of academic intelligence models. Consequently, empirical discoveries about the structure of human abilities are easier to interpret. Academic intelligence research has already proven the theoretical and empirical significance of faceted models. For example, content-related ability factors are contained in Guilford's SOI and in Jäger's BIS-Model (Guilford, 1967; Jäger, 1982, 1984; see Chapter 4.1 for the benefits of faceted models of intelligence). According to Cattell (1987), "concrete discoveries will take on their due richness and meaning only when they are sifted and placed in perspective of classification" (p. 61). Taxonomic foundations can serve several purposes. From a theoretical perspective, they help to differentiate structure and extend existing theoretical models, especially of supposedly heterogeneous constructs, and may provide the basis for a faceted model of intelligence. From a methodological viewpoint, the taxonomy can be used for the construction of new and for the allocation of already existing tests. For existing tests and subtests, initially unstructured and confounded variance sources will be disentangled so that the pattern of covariance can be interpreted more profoundly. Thus, method-related variance related to the different elements of the taxonomic elements can possibly be balanced. When applied during test development, the representativeness of task material and thus, the content validity of the test can be enhanced.

Besides the classificatory principles described in the context of the faceted models of Guilford (1967) and Jäger (1982, 1984), two further taxonomic approaches are apparent in literature, one in the context of academic intelligence models, one in the context of interpersonal perception.

a) Cattell (1987) introduced a theoretical schema of ability dimensions that differentiated between *ability actions* (e.g., involvement of input information in perceptual abilities, involvement of storage and processing components in memory abilities, etc.), *ability contents* (i.e., contents provided by cultural dimensions and contents classified according to the usage of different physiological channels), and *ability processes* (e.g., demands on the ability in terms of the complexity, amount of retentive and retrieval activities, amount of speed activities, etc.). Contents provided by the cultural dimensions are, for example, "verbal (semantic), numerical, spatial, and mechanical contents, *social contents*, arts, music, and science" (p. 72). Contents that enter processing via different physiological channels were subdivided, for example, into visual, auditory, kinesthetic, tactile, and motor contents. Comparable to Guilford (1967), Cattell's (1987) classification sees a social content domain as independent from contents that are related to the type of cue representation.

b) Cline (1964) classified existing measurement approaches to interpersonal perception (i.e., social understanding) according to the stimulus information (e.g., photographs, motion pictures, live behavior, tape recordings, test scores, written material, etc.), the types of instruments (e.g., trait-rating procedures, postdiction of real life behavior or test or item scores, etc.), the sources of the criterion information (e.g., self-provided information, group responses, associates, or experts), and the scoring procedures (e.g., number of correct predictions, difference scores, correlation statistic, quantified evaluations of open responses, etc.). Some of these differentiations are relevant rather for a classification of measurement approaches, but some will be included in the subsequent considerations.

So far, the performance model of social intelligence differentiated only between the cognitive operations (i.e., understanding, memory, perception, creativity, and knowledge). Existing definitions of social intelligence (see Table 4.1) already provide an approximation for some important further distinctions. Some taxonomic principles will be derived from models of academic intelligence. Moreover, taxonomies and empirical results from social psychological research will be introduced in the upcoming chapters. The taxonomic principles addressed are: process variables, outputs, contents and cues, contexts, and targets.

4.3.3.1 Process Variables

The present passage is concerned with a more a profound look into the cognitive operations that constitute social intelligence.

Support for the Performance Model from Social Cognitive Research

Bless, Fiedler, and Strack (2004) presented a comprehensive overview elaborating the present state-of-the-art research on social cognition. Included within this review was a section detailing their concern over the differentiation of social cognitive functions. They identified *perception and attention*, *encoding, interpretation*, and *storage and retrieval* as separate functions. Comparably, Bernieri (2001) mentioned *(a)* the opportunity to experience (i.e., open communication channels and opportunity to perceive), *(b)* attention, and *(c)* the available capacity devoted to perception and inference processes as necessary conditions for correct social judgments. According to Bless et al. (2004), the *perceptive functions* mainly consist of the pure physiological perception process that lets stimuli enter any further cognitive processing. In the perceptive process, attention plays a crucial role. In real life, the devoting of attention to a stimulus is influenced by the salience of the stimulus provided by the situation or by self-raised thoughts or interests. In transferring this idea into social intelligence research, the instructions for tasks take a central position. Thus, the instructions should produce salience of the stimuli that have to be perceived. Particularly, for social understanding the instructions have to direct attention to the relevant situative and personal cues. If this is not accomplished, performance in social understanding can as well depend on whether or not participants have perceived the relevant cues. Bless et al. (2004) suggested that a later memory test can provide evidence of whether the cues were perceived or not (i.e., if yes, they should be remembered).

The *encoding* functions in the model of Bless et al. (2004) reflect the process of assigning a mental representation to a perceived stimuli (e.g., assign “smile” to a perceived movement of a person’s mouth, or “man” to a perceived person). In general, available social categories are used to encode stimuli into mental representations. Within the social intelligence framework, this cognitive process can also be subsumed under the social perception abilities since Bless et al. still exclude any interpretative demands. In contrast, *interpretation* reflects any further information processing that goes beyond the given information. Analogous to the model of social intelligence, *storage and retrieval functions* represent social memory and social creativity, respectively. Bless and colleagues reported studies that varied the instruction to a memory task about behavior descriptions of people. One study directly instructed participants to memorize and recall the provided information, while the other instructed them to form an impression of the described person. Participants of the latter condition showed better performance. According to the conclusions of the authors, storage of information thus depends on the depth of information processing.

Interestingly, personal knowledge is supposed to influence all other social cognitive functions. For example, attention is devoted to a person stimulus when it diverges from what the perceivers know about the person. Since social knowledge is organized into categories, it also influences the encoding of information. Moreover, stimuli can more easily be stored when they can be integrated into a knowledge network. Finally, the interpretation of a stimulus depends on the personal knowledge about the person or the situation. In summary, this classification of social cognitive functions support the relevance of the operative ability domains of social intelligence as established in the modified performance model of Weis and Süß (2005), acknowledging also the special role played by social knowledge.

Social Understanding – One or Multiple Processes?

The upcoming passages will only focus on the cognitive processes and functions underlying the ability of social understanding. In particular, social psychological research was concerned with dissecting the process into various dimensions so that it is frequently doubted whether the ability to judge others is a unitary ability (Bronfenbrenner et al., 1958; Cline, 1964; Gage & Cronbach, 1955). One differentiation opposed automatic and controlled information processing (Bless et al., 2004), which is similar to Probst's (1982) distinction of a so-called "Diskursivmodell" versus "Intuitivmodell" (i.e., inference by analysis vs. inference by intuition). Automatic processes are unintentional, require only few resources, and occur in familiar situations. On the contrary, controlled processes require many resources and are applied in unfamiliar situations. Controlled processes demand the deliberate sampling and analyzing of available social cues, whereas automatic processes typically make use of mental shortcuts such as *(a)* the empathic transposing into the role of another person, *(b)* putting the situation, the target's behavior, or the sequence of events in relation to what the judge has already experienced, or *(c)* an imitation of the other person's movements in order to create the same inner sensory information (Bless et al., 2004; Tagiuri, 1969). With respect to the last point, an individual's own mental state can guide further information processing (e.g., "How do I feel about it?") and lead to a conclusion about the mental state of the target.

Bless et al. (2004) applied the distinction of bottom-up versus top-down processing. This reflects the relative impact of new stimuli and prior knowledge on the inference process. Comparably, Buck (1983) disentangled the cognitive process of social understanding into perception and knowing. In both models, direct perception (or, bottom-up processing, data-driven) represents the exclusive use of the available stimuli in order to make an accurate judgment (i.e., perception of the relevant cues as a necessary and sufficient condition). Top-down processing (concept-driven) still relies on the stimuli but includes further information processing based on knowledge. The selection of an adequate processing model in a specific situation depends on the *processing capacity* and *processing motivation* of an individual (Bless et al., 2004). Bottom-up processing is highly resource-consuming so that a restriction of resources should instead evoke top-down processing. Furthermore, the processing motivation (i.e., strive for accuracy) is supposed to be enhanced when the individual is involved in the situation. This can hardly be accomplished in group testing situations, therefore, the instruction should enforce this strive for accuracy.

In a different approach, Bronfenbrenner et al. (1958) distinguished between the *sensitivity to the generalized other* and *the sensitivity to the individual*. According to Bronfenbrenner and colleagues, it addresses independent ability domains that are combined additively in social judgments. Moreover, the distinction also reflects a classification of targets of judgments. The *sensitivity to the generalized other* represents the ability to identify characteristics of what people have in common and requires knowledge of the social norm or the typical (average) response of the respective group (stereotype accuracy). The *sensitivity to the individual* represents the ability to recognize when and how much individuals diverge from a group average (i.e., the individual's emotions, motives, thoughts, etc.). Bronfenbrenner and colleagues did not state whether individuals explicitly distinguish between both processes. However, it can be assumed that they are seldom applied explicitly.

In summary, the paramount concern refers to the dimensionality of the ability domain social understanding. At present, social intelligence research did not systematically account for the questions of the underlying processes when different social understanding tasks are accomplished. Some more concrete considerations that imply the aforementioned models are presented in Chapter 5.2.4 where the scoring of social understanding tasks is addressed. More detailed considerations are provided in Chapter 10 that point towards interesting future research questions on the aforementioned concerns.

4.3.3.2 Queried Information or Product of Tasks

The queried information represents the output (product) which is derived from the cognitive operations. The SOI already includes a product facet distinguishing between six elements (i.e., units, relations, systems, classes, transformations, and implications). For example, units represent discrete social cues such as facial expressions, relations stand for the social-emotional relationship between two people, and systems represent conclusions about the relationship of three or more people (O'Sullivan, 1983, 2007). However, the taxonomic principles of the SOI were criticized for being an arbitrary classification that does not represent the present developments of the social intelligence construct (Cattell, 1987; Probst, 1973). Definitions of social understanding (see Table 4.1) also include statements about the queried information: feelings (e.g., moods, emotions), cognitive components (e.g., thoughts, intentions, attitudes), and more general concepts such as personality traits, future behavior, and relationships to one or more people. Analogously, Bernieri (2001) differentiated between emotions, thoughts, intentions, relationships, and personality traits. Obviously, queried information can be classified into modality categories. These modalities refer to the psychological qualities *inside* (i.e., a person's emotions, cognitions, and personality traits) and *outside* the person (i.e., the person's relationships with other people).

The single categories of the modalities represent broad fields of psychological concepts and are each represented in separate psychological disciplines (i.e., psychology of emotion; cognitive psychology; social psychology; etc.). Here lies a meaningful interface between differential and general psychological approaches. A detailed description of the single disciplines and their results reaches beyond the scope of the present work and is also not relevant. The benefit, however, of using general psychological research results lies in two aspects. *(a)* They can provide further differentiations of the category of cues. These differentiations need to be sufficiently distinct and broad in order to provide a reasonable classification. *(b)* For specific types of (sub)constructs, they can serve as an expert-based database for determining correct answers to items when no objective information is available. An approach which is applied to the latter principle is described in MacCann (2006) and will be addressed in a later section when the measurement instruments are described. At this point, only the first aspects are addressed.

Emotions

With respect to the modality of emotions, a lot of research was directed at the identification of interculturally and universally valid emotions in terms of the patterns of appraisal and expressions (Ekman, 1999; see Scherer, Schorr, & Johnston, 2001 for a comprehensive presentation). Basic emotions can be identified by a common activity pattern of the autonomic or central nervous system (Ekman, 1999). For example, Ekman (1999) identified 15 distinguishable emotions: *amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness / distress, satisfaction, sensory, pleasure,* and *shame*. These universally valid emotions should be detectable in other persons when applied as queried information in a performance test. Whether other emotional experiences must be treated differently when they enter performance testing is not clarified. Another question concerns whether emotional judgments of more than one emotion could be dependent on each other when occurring at the same time with reference to the same situation. The dependency, if it does exist, would be reflected in the personal knowledge about the emotional system. Present tests of emotion recognition in the context of emotional intelligence treat the dimensions as independent (Mayer, Salovey, Caruso, & Sitarenios, 2002).

Cognitions

The modality of *cognition* even seems more complex than that of emotions. Definitions of social intelligence speak of thoughts, intentions, attitudes, motives, etc. Ford (1992) conceived intentions or goals as the anchors that organize the activities within a behavioral episode. This represents an organized pattern of cognitive, emotional, biological, and perceptual-motor activity. Therefore, the goals are of special interest for finding thematic classifications of the contents of the cognitive elements that guide behavior.

Ford (1992) distinguished between *directive cognitions* (i.e., personal goals such as a desired outcome of a situation), *regulatory cognitions* (i.e., evaluative functions that determine whether a goal or an activity is of a certain value), and *control cognitions* (i.e., planning and problem-solving thoughts). Within the personal goals, Ford applied the Ford and Nichols Taxonomy of Human Goals (Ford & Nichols, 1987 as cited in Ford, 1992) as a comprehensive taxonomy of goal contents for a better understanding of the direction and organization of human behavior. In the taxonomy, *affective goals, cognitive goals, subjective organization goals, self-assertive social relationship goals, integrative social relationship goals*, and *task goals* are differentiated and elaborated upon. The Ford and Nichols Taxonomy further distinguishes between several elements within one goal domain which will not be addressed here.

There are surely other taxonomies of cognitions and goals available in the literature (see Ford, 1992; see also Bless et al., 2004 for an overview from social cognitive research) that will not be addressed any further.

Relationships

Relationships can also be described in terms of various dimensions. For example, Tagiuri (1969) introduced the dimensions of friendship, love, power, and influence as descriptive qualities for interpersonal relationships. A more theory-guided classification is provided by the interpersonal circumplex model (Wiggins, 1979). The interpersonal circumplex is a nomological system that serves different purposes. It classifies interpersonal personality traits in terms of their conceptual similarity and serves as a validating framework for the construct validity of interpersonally relevant traits. It includes four bipolar dimensions arranged in a circle which are given different labels throughout different approaches. The two core dimensions are labeled *love* (*distance vs. closeness*) and *power* (*dominance vs. submission*); these two dimensions are orthogonal (Wiggins, 1979). In between, two further dimensions are located that are again orthogonal. The dimension of *competition* includes the poles *competitive* vs. *cooperative*, with the competition pole located in between dominance and distance. The dimension *agency* with the pole of *extraversion / activity* (vs. *introversion / passiveness*) is located in between dominance and closeness. The single dimensions can also be applied to describe the character of interpersonal relationships in general or with respect to one person (e.g., How dominant is someone in his or her social relationships in general? How dominant is someone in his or her relationship to his or her partner?).

Personality traits

Personality psychology provides numerous classificatory systems of personality traits: the Big Five (Angleitner, Ostendorf, & John, 1990; Costa & McCrae, 1992; Goldberg, 1990), Eysenck's three factor model (Eysenck & Eysenck, 1969), and Cattell's 16 Personality Factors (Cattell, 1950). They shall not be addressed any further. Suffice to say that all the established models and the respective questionnaire inventories can serve as a basis for determining the personality traits of targets and to let judges answer questions about these respective traits.

4.3.3.3 Cues and Contents

Cues: Person and Situation

This section refers to the cues used to perform any social cognitive task. The cues represent the data basis for any information processing, be it in social, emotional, or academic intelligence. Cues determine the output of a cognitive process to a substantial extent besides the knowledge structure of a judge that exists independently from the cues. Cues can be provided by a person (i.e., the target) and by a situation (Tagiuri, 1969). Situative cues can reduce the number of possible explanations for a set of person-related stimuli. Tagiuri concluded that both the person and the situation separately from one another allow "nonrandom, but indeterminate, judgments" (p. 421). Highly determinate judgments should be reached when both sources provide information.

From the perspective of a judge, cues can also stem from inside the person (e.g., somatosensory information, feelings, emotions, etc.). Especially for the construct of emotional intelligence, cues from inside oneself make up a meaningful conceptual domain that will be discussed in Chapter 4.3. Cues vary in terms of the unambiguousness for the respective task. Univocal cues require less further information processing to solve a task. Univocity can only be determined with respect to an underlying task, the context, and last but not least, the features of a stimulus. Empirical studies generally support the importance of the context for the perception and interpretation of person-related cues (e.g., Gestalt theorists) (Archer & Akert, 1980; Bless et al., 2004). Thus, the quantitative and qualitative properties of the situation and the stimuli must not be ignored.

Archer and Akert (1980) addressed the question of the type and amount of information (i.e., cues) necessary to accomplish a social understanding task. They had previously discovered that a high level of accuracy was reached although only limited information was provided. Archer and Akert established three theories of social interaction that provide different explanations for the relationship between the availability of information and the accuracy of judgment. The *Additive Theory* states that accuracy accumulates as a linear function along with a growing amount of information or different information channels. The *Significant Clue Theory* postulates that indispensable information is highly localized in terms of place and time during the stream of information. Stimuli in measurement instruments or in real-life frequently contain controversial or misleading information which should not be taken into account. Thus, it is necessary to differentiate between 'good' and 'bad' information. The *Diffusion Theory* claims that any individual piece of information provides sufficient information for accurate performance. In several investigations, Archer and Akert systematically manipulated the available information. They varied the length of the stimuli and the breadth of communication channels and assessed the accuracy of social understanding. Results showed that, in any event, the availability of the full information (in length and breadth) yielded the best performances across all types of scenes and material. However, in some cases, the performance reached a comparable level when only one part of a scene or just one piece of information was provided and nearly every condition reached an above-chance level of performance. But accuracy varied substantially across and within the scenes (i.e., material-dependent) so that the interpretation of the results was equivocal. The authors carefully interpreted their findings in favor of the Significant Clue Theory in combination with the Diffusion Theory.

Cues may also vary according to their relevance for a task and for the underlying ability construct. Mehrabian & Ferris (1967) investigated the relevance of different communication channels. They found that more than half of communication relies on body language (55 %; e.g., postures and gestures) while 38 % are based on cues from spoken language (e.g., the tone of voice), and only about 7 % of the communication relies on the content of the language. Ekman, Friesen, O'Sullivan, and Scherer (1980) investigated the effect of different cues (i.e., face, body, speech, and a combination of all) on the judgments of personality characteristics. However, judgmental accuracy was not assessed. Instead, just the correlations between the judgments based only on the specific cues and the judgments based on a combination of all of the cues was assessed. Results showed that the relevance of the specific cues varied according to the judged attributes (i.e., the two judgments converged to different extents for different types of cues). In contrast to Mehrabian and Ferris (1967), Ekman et al. (1980) conclude that no communication channel is more or less important than any other, but that it is rather the interaction of the cues, judged attribute, and situation that is significant.

How Can Social Cues Be Presented?

The presentation of social cues or information and the applied task material (i.e., task contents) are directly associated with each other. Some models of academic intelligence already include a differentiation of different content domains, for example, symbolic, semantic, figural (i.e., visual and auditory), and behavioral contents in the SOI (Guilford, 1967) and verbal, numeric, and figural-spatial contents in the BIS-Model (Jäger, 1982). The SOI is not an appropriate guideline for classifying content domains of social intelligence (see also Probst, 1973) since it abandons important components from the behavioral contents because auditory contents (e.g., the tone of voice or the way of speaking) are seen as a separate content domain. Cattell (1987) also criticized the arbitrary selection of content domains in the SOI.

Definitions of social intelligence include *words* and *facial expressions* as possible social cues besides the aforementioned cues contained in spoken language. O'Sullivan et al. (1965) distinguished between facial expressions, vocal inflections, postures, and gestures as cues from which mental states are inferred. Consequently, relevant social cues can be contained in written language (e.g., the way people use the language in order to express an emotion, etc.), in audio recordings of spoken language (e.g., tone of voice, way of speaking, language contents), and in presentations of the person's face and body (e.g., facial expressions, body language, etc.). Typically, situative contents can be presented in the same type of material.

4.3.3.4 Settings

The differentiation according to the setting refers to the situative characteristics that may influence information processing. For example, settings may be located in private contexts (e.g., family and friends) or public contexts (including professional settings or encounters with official organizations). It is possible to assume that people who score high on social intelligence in public contexts may show low social intelligence in private settings. Another classification accounts for the varying numbers of people involved in the situation (i.e., a person is alone, in a dyadic interaction, or member of a group of people). People may feel more comfortable with less people involved, which results in them thinking and acting differently. Other aspects concern the psychological qualities of the situation (e.g., is the general atmosphere a pleasant one). Possible taxonomic principles to classify these psychological qualities are again found in the interpersonal circumplex dimensions (Wiggins, 1979). In this respect, the core bipolar dimensions of the interpersonal circumplex (i.e., love and power) can be applied to classify to underlying theme of a situation (e.g., when two people are having a fight about a major job-related decision, this situation would be classified to the power dimension of the interpersonal circumplex).

4.3.3.5 Targets of Judgment

Definitions of social intelligence distinguish between *oneself* and *others* as the targets of social cognitive operations. This refers to the question of who shall be judged in terms of his or her mental states (e.g., the interpretation of a stranger's behavior for judging his or her intentions; remembering another person's physical appearance so that he or she is recognized at a later point in time). The relevance of applying socially intelligent operations to *oneself* is not further discussed in the social intelligence literature. However, the ability to understand one's own emotions and cognitions, to remember relevant autobiographic elements or past experiences, or to perceive one's own behavior is definitely relevant for showing socially intelligent behavior. Emotional intelligence models address this question in more detail (see Chapter 4.4.1).

The category *others* can be further subdivided. Bronfenbrenner et al. (1958) distinguished between first-, second-, and third-person sensitivity. *First-person sensitivity* refers to the question of what the judge thinks about how others think and feel about him or herself. *Second-person sensitivity* asks the judge to make a statement about how the target thinks and feels about the target him or herself. *Third-person sensitivity* addresses the judgment of what the target thinks and feels about someone or something else. Definitions of social understanding introduce *strangers* as a special case of the category *others*. This refers to the relationship of the judge to the targets of judgment in terms of familiarity. This dimension can be assumed as a continuum with the extremes of completely unfamiliar (i.e., strangers) versus very familiar. Familiarity can be expressed in terms of qualitative and quantitative criteria. Literature shows that performance in social cognitive tasks varies with familiarity in a positive direction (Herzmann, Danthiir, Wilhelm, Sommer, & Schacht, 2007; see Buck, 1983 for an overview of results). Herzmann et al. (2007) reported empirical studies showing an increase in accuracy and a decrease in response latencies in a task of emotion recognition when familiar faces were applied as stimuli. Growing familiarity can also result in qualitative differences of the ability demands because the amount of previous knowledge about the target is richer and information processing may increasingly be based on this knowledge. In this respect, Sabatelli, Buck, and Dreyer (1980, 1982) could, in fact, not find a relationship of the general level of accuracy to interpret the nonverbal cues of partners with the general marital satisfaction across all cues. However, the accuracy was positively related to marital satisfaction when only those cues were accounted for, which could not be accurately interpreted by unfamiliar people. Buck concluded that these finding seem to limit the possibility to construct general measures of nonverbal understanding when the measures are based on unknown persons expressing cues that are highly idiosyncratic and specific to the sender.

Another issue dealing with the judge-target-relationship is the interaction of common and uncommon characteristics in both (i.e., the familiarity of judge and target). Cronbach (1955) distinguished between an assumed and real similarity. He postulated better performance when assumed and real similarity are either both high or both low (see also Chapter 5.2.4). Bruner and Tagiuri (1954) reported positive correlations of the assumed similarity between a judge and target and the accuracy of those judgments. Bronfenbrenner et al. (1958) found that male judges' accuracy in judging other males correlated positively with their accuracy in judging women. Contrary, females' accuracy in judging other females correlated negatively with their accuracy to judge other males. The effect of assumed and real similarity on the accuracy of judgments is especially problematic when only one target has to be judged, which results in the influence being unbalanced across tasks.

4.3.4 Social and Academic Intelligence – The Same "*g*"s, Domain-Specific Overlap, or Independent Intelligences?

Only Guilford's (1967) Structure of Intellect Model included a statement about the construct overlap of academic and social intelligence. In the SOI, he represented coequal ability domains. The taxonomic weaknesses of the SOI have already been discussed (Cattell, 1987). Accepting Guilford's model would mean that social intelligence shares everything with academic intelligence (i.e., cognitive operations and task products) except for the task contents. It would also mean that auditory, figural, or semantic task contents are not relevant for social intelligence. Neither other theoretical accounts nor empirical studies suggested a coherent model of overlap between social and academic intelligence (see Chapter 5.4 for the results of validity studies of social intelligence). In the end, this question is an empirical one, however, construct definitions suggest that there are more than one hypothetical models of overlap.

Conceptually, the largest overlap of the modified performance model of social intelligence can be identified in the faceted BIS-Model. Overlap can be determined from different perspectives related to the facets. Theory suggests that social and academic intelligence share common operative and content-related components. The operative domains *reasoning*, *memory*, *perception (speed)*, *creativity*, and *knowledge* are all included in models of both constructs. At the same time, verbal and nonverbal (figural-spatial in the BIS and picture-, and video-based in the social intelligence model) content components are included in both models. Auditory task contents are, so far, only included in the model of social intelligence. Consequently, in order to fully validate social intelligence against the BIS-Model, this must be complemented by an auditory ability domain. The doctoral thesis of Kristin Seidel (2007) extended the BIS-Model by newly developed auditory ability tasks. Results showed that a differentiation into operative domains (i.e. auditory reasoning, memory, etc.) was not possible. Tasks showed the highest correlations with the reasoning domain. With respect to the task contents, a differentiation into language-based and nonverbal (i.e., based on tones) tasks was supported by data. These findings at first contradict the idea of a separate auditory content factor since the identified factor is confounded with common reasoning requirements. However, Seidel pointed out that further research is needed. Thus far, the present work will apply the BIS as a model of reference without an additional content ability domain.

The overlap between social and academic intelligence for the example in the BIS - Model, could be described as displayed in Figure 4.6 *(a)*. The faceted structure of the BIS-Model is complemented either by an additional social-operative or social-content domain. This is an example of integrating social and academic intelligence in one theoretical model. Both classifications are possible in the terminology of the performance model of social intelligence by Weis & Süß (2005). A necessary condition for both types of classification is the identification of and empirical support for a structural model of the respective operative or content-related ability domains of social intelligence. If social intelligence can be partitioned into meaningful content-related ability factors (e.g., written or spoken language, picture- or video-based) subsuming under the content factors in the BIS-Model, social intelligence may represent an additional operative domain. However, this classification is somehow difficult. The differentiation of contents is not fully equivalent, numeric contents are missing in social intelligence, and no auditory contents are, thus far, incorporated in the BIS-Model. In contrast, conforming to the previously described performance model of social intelligence, social abilities may more suitably represent an additional content domain if splitting into the operative ability factors (analogous to the SOI). No matter what overlap emerges from this faceted perspective, the possibility is maintained that social and academic intelligence subsume under one general factor. Another type of visualization for these hypotheses is displayed in Chapter 6.3 (Figure 6.4) as a hypothetical model for conducting confirmatory factor analysis (CFA) to examine construct validity.

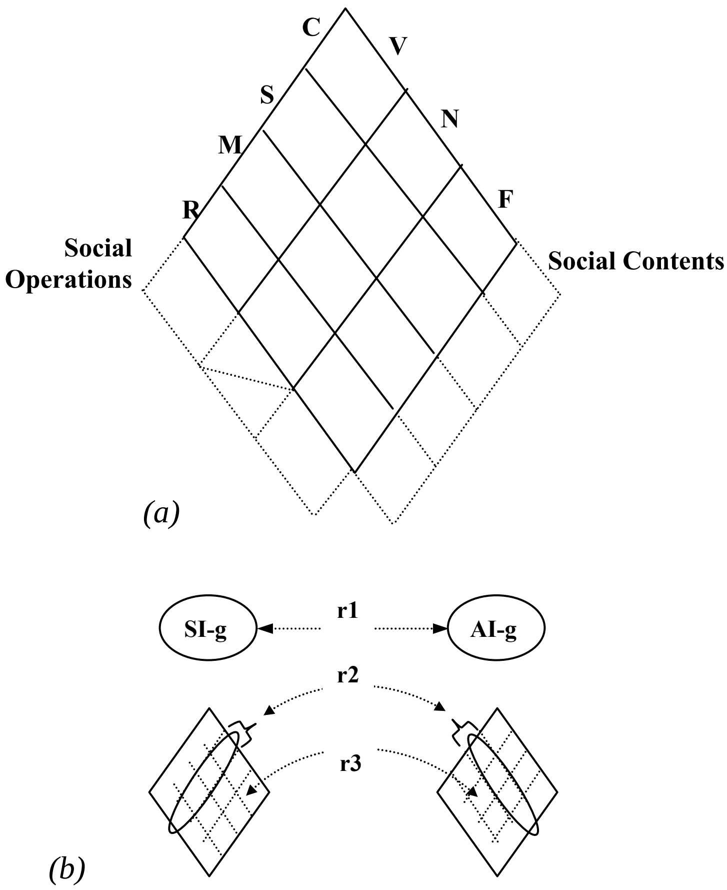


Figure 4.6

Overlap of Social and Academic Intelligence (a) On a Structural Level Based on Common Facets, (b) On Different Levels of Generality

Note. R = Reasoning, M = Memory, F = Figural, C = Creativity, S = Speed, V = Verbal abilities, F = Figural abilities, N = Numerical abilities
r1 = relationship of social to academic intelligence on a general level, r2 = relationship of social to academic intelligence on the level of broad ability factors, r3 = relationship of social to academic intelligence on the lowest level in the hierarchy

Figure 4.6 *(b)* displays a different model of construct overlap that turns relevant when both constructs are specified as a hierarchical model. Consequently, overlap cannot be restricted to just one ability facet (operations *or* contents). In contrast, it can occur on different levels of generality when a hierarchical model is assumed on both sides (i.e., positive correlations *r1* between two *g*-factors, between corresponding broad ability domains *r2*, or between single cells *r3*; see Figure 4.6 *(b)*). In any case, whether social intelligence is distinct from academic intelligence can only be determined empirically.

4.4 Ability Constructs Related to Social Intelligence

At the beginning of the present work, emotional intelligence was introduced as a competitor to social intelligence as a valuable new intelligence constructs that is sought to complement the traditional academic intelligence construct. Besides emotional intelligence, practical intelligence was introduced also not long ago to compete with academic intelligence. The concept of wisdom is not a typical human ability construct but originates from research on life-span development. All three constructs show substantial conceptual overlap with social intelligence. Sternberg's (1997) concept of successful intelligence or Gardner's (1983) conception of multiple intelligences were also mentioned in the context of new ability constructs and show some commonalities with social intelligence. However, they will not be accounted for in the present description. One reason for this decision is the lack of empirical support for these concepts and their overinclusiveness in terms of their applied construct definitions (Matthews et al. 2005). Furthermore, emotional or practical intelligence already cover some of the subconcepts of a successful intelligence in Gardner's multiple intelligences.

4.4.1 Emotional Intelligence

Salovey and Mayer (1990) introduced emotional intelligence in psychological literature. They defined emotional intelligence as a subset of social intelligence that involved "the ability to monitor one's own and other's feelings and emotions, to discriminate among them, and to use this information to guide one's thinking and actions" (p. 189). With their conceptualization, they attempted to overcome the sometimes promoted dualism of the emotional and the cognitive system by putting emotional intelligence at the intersection between emotion and cognition (Forgas, 2000; Matthews et al., 2005; Mayer & Salovey, 1997). Traditional positions assumed that affect reduced individuals' ability to think rationally. More recent research conceives "affective states as a meaningful source for the regulation of cognitive processes" (Bless et al., 2004, p. 179).

In the year 1995, Goleman published his notorious book "EQ – Why it can matter more than IQ" and initiated a surge of research which accounted for a simultaneously growing market for the assessment and training of so-called soft skills and for emotional intelligence as a prototypical representative. Throughout subsequent research, the fathers of the construct Mayer et al. (2000) distanced themselves from those strands of research that presented emotional intelligence as an overinclusive new ability construct or which conceptualized emotional intelligence rather as a personality trait. Particularly, these approaches were criticized for being overinclusive by incorporating personality traits, behavioral skills, as well as emotional competencies or, in other words, everything else but IQ (Hedlund & Sternberg, 2000). Mayer et al. (2000) labeled these approaches mixed models or trait emotional intelligence (see also Petrides & Furnham, 2001) in contrast to ability models representing a mental ability. This distinction was adopted by subsequent research.

Figure 2.1 in the introduction chapter displays the development of publications throughout the last 47 years. The number exploded for emotional intelligence and has now nearly reached the level of social competence as a highly general and established construct both, above all, in the field of applied psychology. It has outperformed research on social intelligence within less than 10 years. Approaches to conceptualize and assess emotional intelligence are as numerous as manifold and it is not possible to present an exhaustive description in the present work. Instead, the prototypical approaches will be presented.

4.4.1.1 Ability Models of Emotional Intelligence

Four-Branch-Model of Emotional Intelligence (Mayer & Salovey, 1997)

The current state-of-the-art in terms of theoretical models, is represented by the Four-Branch-Model of Emotional Intelligence (Mayer & Salovey, 1997). The Branches contain emotion-related abilities that are classified according to the complexity of the cognitive requirements and rely upon each other (Mayer et al., 2000). Figure 4.7 shows the Four Branches arranged from bottom to top.

Branch I (Perception, Appraisal, and Expression of Emotion) involves the most basic requirements of perceiving, recognizing, identifying, or expressing emotional information in one’s physical states, feelings, thoughts, in other people and in artwork. *Branch II (Emotional Facilitation of Thoughts)* represents the ability to assimilate emotions in order to enhance intellectual functioning, for example in order to support judgment and memory. *Branch III (Emotional Understanding)* includes the ability to reason about emotions and to understand transitions from one emotion to another. Finally, *Branch IV (Regulation of Emotion)* refers to the ability to manage emotions in oneself and in others in order to promote emotional and intellectual growth.

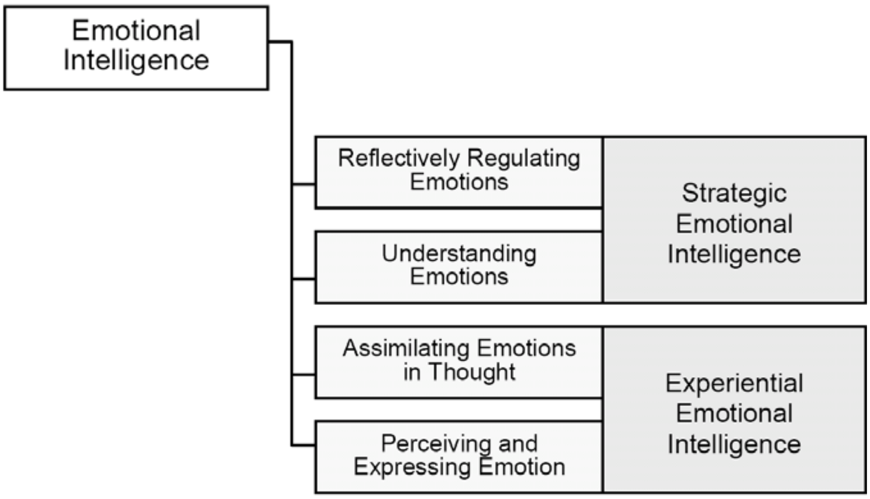


Figure 4.7

Four-Branch-Model of Emotional Intelligence (Mayer & Salovey, 1997; Adapted from Mayer, Caruso, & Salovey, 1999, p. 269)

Mayer and Salovey (1997) postulated positive relations between the Branches and, thus, a hierarchical model. They sought to conceptualize emotional intelligence as a mental ability and expected positive relations to other mental abilities as required by Austin and Saklofske (2005; see also Schaie, 2001). Furthermore, Mayer, Salovey, Caruso, and Sitarenios (2001) postulated a positive relation between performance and age and experience. Particularly, they claimed that the processing of emotional information as specified in Branch III (Emotional Understanding) is based on acquired emotional knowledge that typically covaries with age (e.g., emotional knowledge may contain the claim that anger typically goes together with the experience of injustice). The Mayer-Salovey-Caruso-Emotional-Intelligence-Test (MSCEIT; Mayer et al., 2002) was developed to operationalize the Four-Branch-Model. This test will be presented in a subsequent chapter. The Four-Branch-Model is surrounded by controversy; some of the critical points are discussed hereafter.

a) Theoretical Validation

The only empirical support for the entire model is based on the MSCEIT. If the model was valid, different measures that match the requirements of the Four Branches should also support the postulated structure of emotional intelligence. At the moment, empirical research is only concerned with finding adequate measurement procedures. Hardly any attempts are being undertaken to advance the theory. However, this would be a necessary step in the light of the subsequently presented criticisms.

b) Emotions vs. cognitions

The model fails to explicate the interplay of the emotional and the cognitive system that were sometimes conceived as opponents (Mayer et al., 2001). At first sight, it seems to be a contradiction to speak of emotions within a construct that is supposed to be a cognitive intelligence construct (Carroll, 1993; Süß, 2001). In order to clarify the concepts, one distinction is indispensable: cognitions and emotions can be seen as a process (e.g., the process of thinking, the process of emotion appraisal); they can also be seen as contents that the human mind deals with (e.g., thinking about how it must be to feel depressed, an emotion is elicited by a sudden thought that comes to one's mind). Such an explicit distinction is not realized in the Four-Branch-Model which is equivocal about the question of whether the emotional system also exhibits more or less intelligent processes. Relying on the classical definition of intelligence, only cognitive processes or operations are allowed. Consequently, emotional intelligence could be defined as the ability to apply cognitive operations to emotional contents. Emotional contents can be provided by external cues (i.e., someone speaks about his feelings, or shows body language that reveals information about the emotional life) or raised by emotional processes of one's own emotional system (i.e., appraisal of emotion in oneself as reaction to external cues). However, some Branch definitions do not explicitly clarify the role of emotional processes and contents and suggest an undefined interplay of emotional and cognitive processes. Thus, the Four-Branch-Model obviously lacks taxonomic principles that clarify the operative and content domains of emotional intelligence.

Table 4.2 attempts to disentangle the processes and contents of emotional intelligence as specified in the Four-Branch-Model. It opposes emotionally and socially intelligent operations and contents. Note that not every subcomponent of the Four Branches can be univocally classified to the present taxonomy: *Perception of emotions in the self* can be an automatic process with emotions raised by emotional processes or by a more conscious cognitive process which only includes emotions as contents. Moreover, typical emotion-related traits may influence the *perception and the regulation of emotions in the self* (Gohm, 2003; e.g., people high in impulsivity or low in emotional stability do not encounter the same performance conditions as people with a disparate trait profile).

Table 4.2

Cross-Classification of Processes and Contents of Emotional Intelligence

<table>
<tr><td colspan="2"></td><th colspan="4">Process</th></tr>
<tr><td colspan="2"></td><th colspan="3">Cognition</th><th>Emotion</th></tr>
<tr><td rowspan="2">Contents</td><th>Cognition</th><td rowspan="2">Social Intelligence (Under-standing, Memory, Perception, Creativity)</td><td></td><td colspan="2">Branch II: Emotional Facilitation of Thought
Branch IV: Regulation of Emotions</td></tr>
<tr><th>Emotion</th><td>Branch I: Emotion Perception in others (limitedly in the self)</td><td>Branch III: Emotional Understanding</td><td>Branch I: Appraisal of Emotions in the self</td></tr>
</table>

Besides an unclear structure of performance when *oneself* is involved as the target of perception or regulation, *appraisal and regulation of emotions in others* contain behavioral requirements that do not match the present taxonomy. Table 4.2 shows that not only cognitive processes are involved in the current conceptualization of the Four-Branch-Model (i.e., above all in Branch II and IV). Consequently, models of emotional intelligence need to clarify the role of emotional processes. On the one hand, they reflect the contents of cognitive operations (e.g., apply cognitive operations that influence the emotional reactions; think about something that alters the present mood). If this is true, the aforementioned incompatibility of the emotional and the cognitive system would be irrelevant since the emotional system would not be involved. On the other hand, could the emotional system possibly be the active system that is responsible for the regulation and management of emotional experiences? Or, is it the interaction of the cognitive and emotional processes that represent emotional intelligence? At the moment, emotional intelligence research and especially the Four-Branch-Model are not capable to answer these questions.

c) Performance requirements

Branch IV and parts of Branch I include behavioral requirements. The ability to *regulate emotions in others* (Branch IV) as well as the ability to *express emotions* or *appraise emotions in others* (Branch I) demand overt behavior. According to Mayer et al. (2001), Branch IV even represents an interface between the cognitive system and personality traits. This conceptualization substantially questions the maturity of the theoretical concepts underlying the Four-Branch-Model. Neither personality traits nor behavioral components should be included in an intelligence construct.

Mayer et al. (2001) claimed that emotional knowledge is a necessary requirement for the accomplishment of tasks of *perception* and *appraisal of emotions* and *emotional understanding*. They leave it unclear to what extent perception, understanding, or knowledge contribute to performance in these Branches. However, the separate contribution can vary to a substantial amount. Even more so, Scherer (2007) demanded the exclusion of emotional knowledge from models of emotional intelligence. Exemplary for a prototypical task, emotional knowledge would assume anger or frustration to be the emotional reaction in the experience of injustice. However, it is possible that individuals differ in how they react during this experience. In this case, emotional knowledge of a prototypical reaction could result in a wrong answer. Instead, perception or understanding could provide the correct answer when the relevant cues are available. The present conceptualization does not allow for an explicit separation of crystallized and fluid components of emotional intelligence. Some researchers speculated that emotional intelligence could represent an additional factor amongst the crystallized abilities in Cattell's theory suggesting that it could be acquired knowledge in specific domains (Davies, Stankov, & Roberts, 1998; Neubauer & Freudenthaler, 2005; Zeidner et al., 2001).

Scherer's Model of Emotional Competence (Scherer, 2007)

Another theory-driven approach was just recently established by Scherer (2007). In criticizing existing ability and mixed models and particularly the use of the term "intelligence", he suggested a model of emotional competence based on his Componential Emotion Theory (Scherer, 2001). He applied the differentiation of emotional mechanisms (i.e., processes) and emotional contents. Nevertheless, he identified the lack of agreement on the nature of the emotional mechanisms as the reason for the disagreement about emotional intelligence or competence.

Scherer (2007) conceives high emotional competence as an optimal functioning of the emotional mechanisms with respect to the domains of *emotion production* and *emotion perception*. Therefore, he clearly focuses on emotions as process whereby emotion perception implicitly contains emotional contents as well. Scherer defines *production competence* as the "production of the most appropriate emotional reaction to different types of events based on adequate appraisal of internal goal states, coping potential, and the probable consequences of events, [... as the ...] adaptive regulation of one's emotional states, both with respect to internal set points and according to the sociocultural and situational context [... and as the ...] efficient emotional communication in social interaction through appropriate expression of one's own state." (p. 107). Thus, production competence includes appraisal, regulation, and communication competences, and is strictly restricted to emotions in oneself. *Perception competence* requires accurate signal perception and recognition (i.e., receiving ability) and is directed at the perception and interpretations of emotions in others. Table 4.3 displays the emotional competence model including the two postulated domains and their different subcomponents. Furthermore, the construct overlap with the Four-Branch-Model is presented.

Table 4.3

Scherer's Model of Emotional Competence (2007) Opposed to the Four-Branch-Model

Competence domain	**Subcomponents (Scherer, 2007)**	**Four-Branch-Model (Mayer & Salovey, 1997)**
Production Competence	Appraisal Competence: *appropriate emotion elicitation and differentiation*	Branch I (Appraisal in the self)
	Regulation Competence: *correct inappropriate emotional responses produced by unrealistic appraisals*	Branch IV (Regulation in the self)
	Communication Competence: - *produce emotional expressions optimally suited to a purpose* - *accurate signal perception and receiving ability* (see Perception Competence)	Branch I (Expression in the self and perception in others)
Perception Competence	*Accurate signal perception*	Branch I (Perception in others)
	Accurate recognition and interpretation	Branch I (Perception in others)

Scherer (2007) additionally introduced different performance criteria for the different competence domains. The so-called Aristotelian model is relevant for the *regulation competence*, whereas the so-called Galtonian model applies to the *appraisal and communication competence* including the domain of perception competence. The Aristotelian model is concerned with the appropriateness of an emotional response (i.e., typically in the middle of a range of opportunities). The Galtonian model refers to perfect accuracy as the desired performance criterion. Obviously, Scherer's model is more restricted in terms of ability domains and range of targets and contents, and, at the same time, substantially more explicit and clear about the necessary performance requirements and the underlying conceptual differentiation. The use of the term *competence* instead of *intelligence* allows a less strict selection of performance requirements. Scherer clearly differentiates between the self and others and minimizes knowledge requirements as specified in Branch III in the Four-Branch-Model. It also excludes those domains in which it could not be clarified whether emotional or cognitive processes contributed to high performance (i.e., particularly Branch II). However, the model requires empirical substantiation foremost, and then, may well serve as a viable alternative to the Four-Branch -Model.

Concurrent / Competing Approaches to Define Emotional Intelligence

Several strands of research have now emerged which provide complementary or competing definitions of more specific ability domains of emotional intelligence (Austin 2004; Freudenthaler & Neubauer, 2005). Austin (2004) focused on a biological perspective and postulated that emotion processing speed is a part of emotional intelligence. Freudenthaler and Neubauer (2005, 2007) focused on the ability domain of *managing emotions* as specified in Branch IV in the Four-Branch-Model. They criticized that existing operationalizations do not cover the effectiveness of subjects' *behavior* in managing emotions, but rather include emotional *knowledge* about the effectiveness of behavior (assessed by a maximum-performance criterion). Alternatively, they suggested a conceptualization of emotional management skills that includes typically exhibited behaviors in everyday life. However, this approach is faced with several conceptual problems relating to the overlap with trait emotional intelligence when the typical performance conditions are applied (see Chapter 5.3). MacCann (2006) relied on the definitions of Understanding Emotions and Emotion Management of Branch III and IV of the Four-Branch-Model. She developed a different measurement approach that will be presented in Chapter 5.3 together with the approaches of Mayer et al. (2002), Freudenthaler and Neubauer (2005), and Austin (2004).

4.4.1.2 Mixed Models of or Trait Emotional Intelligence

Ability and mixed models of emotional intelligence differ both on the construct and on the measurement level. Mixed models literally contain a mixture of ability constructs, personality traits, motivational components and behavioral skills. They are assessed by the use of self-report inventories (Petrides & Furnham, 2001; Saklofske, Austin, & Minski, 2003; Schutte, Malouff, Hall, Haggerty, Cooper, Golden, & Dornheim, 1998). Some authors in this field still claim to assess an intelligence construct. However, intelligence theories and empirical data clearly demonstrated that the use of self-report data for assessing a cognitive ability construct is highly problematic (Paulhus, Lysy, & Yik, 1998; see also Asendorpf, 2002; Neubauer & Freudenthaler, 2005; Scherer, 2007). Goleman (1995) formulated a model of emotional "intelligence" consisting of five major domains (i.e., *knowing one's emotions*, *managing emotions*, *motivating oneself*, *recognizing emotions in others*, and *handling relationships*). Bar-On (1997) defined emotional "intelligence" as an array of noncognitive capabilities, competencies, and skills that influence one's ability to succeed in coping with environmental demands and pressures. His conceptualization contained five domains: *intrapersonal skills*, *interpersonal skills*, *adaptability skills*, *stress-management skills*, and *general mood*. Bar-On's model of emotional "intelligence" is operationalized by the Bar-On Emotional Quotient Inventory (EQ-I; Bar-On, 1997).

Goleman's and Bar-On's models were criticized for being overinclusive (Hedlund & Sternberg, 2000; Matthews et al., 2005). However, Petrides and Furnham (2001) introduced a model of trait emotional intelligence (synonymous for mixed models) that includes 15 dimensions: *adaptability*, *assertiveness*, *emotional appraisal (self and others)*, *emotion expression*, *emotion management (others)*, *emotion regulation*, *impulsiveness*, *relationship skills*, *self-esteem*, *self-motivation*, *social competence*, *stress management*, *trait empathy*, *trait happiness*, and *trait optimism*.

In general, and inclusive to all of the inventories, results from validational studies showed large overlap with measures of personality traits (e.g., alexithymia or emotional stability). Pérez, Petrides, and Furnham (2005) reported moderate to high correlations between personality inventories and 15 selected self-report inventories of trait emotional intelligence. Consequently, the value of mixed models of emotional intelligence is highly disputable in terms of the advancement of theories and the applied methodology. Hereafter, only ability models of emotional intelligence will be taken into account.

4.4.1.3 Overlap of Social and Emotional Intelligence

Several researchers agreed that social and emotional intelligence show conceptual overlap (Davies et al., 1998; Kang, Day, & Meara, 2005; Mayer et al., 2000; Salovey & Mayer, 1990; Weis & Süß, 2005; Weis et al., 2006). Only three empirical studies can be found in the literature that investigated the relation of social and emotional intelligence relying on performance tests (Barchard, 2003; Davies et al., 1998; Weis & Süß, 2007). Davies et al. (1998) found a nonsignificant negative correlation between two performance tests of social and emotional intelligence. Barchard (2003) operationalized social intelligence as a subdomain of emotional intelligence but did not report construct intercorrelations. Weis and Süß (2007) showed that an emotional perception task loaded on a social understanding factor but not on a general social intelligence factor, suggesting domain-specific overlap between the two constructs. However, the empirical evidence is sparse and restricted to single operationalizations. More importantly, common theoretical considerations were not undertaken which would allow large-scale conclusions about the construct overlap (or distinctiveness). Weis et al. (2006; see also Süß et al., 2005) assembled the cognitive operations derived from a requirements' analysis of both social and emotional intelligence tasks. With respect to this analysis and the aforementioned considerations, Table 4.4 contrasts social and emotional intelligence on different taxonomic levels based on the Four-Branch-Model and the performance model of social intelligence.

The operations of *understanding* and *perception* appear in both constructs, being applied to different contents or queried modalities (i.e., emotions for emotional intelligence; emotions, cognitions, behavior, etc. for social intelligence). Thus, social intelligence seems to be the broader construct in terms of the covered modalities. With respect to Branch II and IV, emotional intelligence seems to reach beyond the scope of social intelligence. *Emotional Facilitation of Thought* and *Emotion Regulation* may make people more apt to think and behave intelligently in a specific situation by ruling out possibly unfavorable moods or by self-motivational mechanisms (i.e., meta-cognitions). The contents or cues that social and emotional intelligence are dealing with are totally identical. Social intelligence is only directed at *others* as targets, whereas emotional intelligence focuses both, on the self and on others. Conclusively, measure are needed that vary systematically the aforementioned aspects in order to determine the construct overlap.

Table 4.4

Overlap of Social and Emotional Intelligence

Domain of overlap	Social Intelligence		Emotional Intelligence	Type of Overlap
Operations	Social Understanding	$\subset$	Emotional Understanding (Branch III)	EI: Reasoning requirements applied to emotions
	Social Memory	$\neq$		
	Social Perception	$\subset$	Emotion Perception (Branch I)	EI: Perceptual requirements applied to emotions
	Social Creativity	$\neq$		
		$<$	Emotional Facilitation of Thought (Branch II)	EI: Meta-Ability: Influence on all cognitive processes
		$<$	Emotion Regulation (Branch IV)	
Queried Modality	Emotions	$=$	Emotions	Identical
	Cognitions	$\approx$	(Meta-cognition)	EI: indirect effect on cognitions (Branch II/IV; see Table 4.2)
	Behavior	$\neq$		
	Relationships	$\neq$		
Contents	Written language	$=$	Written language	Identical
	Spoken language	$=$	Spoken language	Identical
	Body language	$=$	Body language	Identical
Targets		$\neq$	Self	
	Other	$=$	Others	Identical

Note. = identical, $\neq$ not equal to, $\approx$ some unspecified overlap, $\subset$ a proper superset of, $<$ Influence on, EI = emotional intelligence

4.4.2 Practical Intelligence

Practical intelligence was frequently conceived as interchangeable with social intelligence (Cantor & Harlowe, 1994; Ford, 1986). Mercer, Gomez-Palacio, and Padilla (1986) equated practical intelligence with social competence and defined the latter as the ability to meet the normative expectations of others in six different types of social roles (i.e., family roles, peer roles, community roles, earner-consumer roles, self-maintenance roles, and non-academic school roles). Thus, the normative expectations are supposed to vary across social roles, cultures, or social groups. Sternberg (1985, 1987, see also Hedlund & Sternberg, 2000) extended the scope of practical intelligence beyond problems of only a social nature onto all types of problems encountered in everyday life. These problems are typically not clearly defined and solutions not readily available. Common to both approaches is the specificity for the respective context or content domain. Thus, practical intelligence is a context-specific capacity defined as the ability to find a more optimal fit between the individual and the demands of the individual's environment by applying information-processing components for the purposes of adaptation to, shaping, and selection of the environment (Sternberg, 1985, 1987). According to Wagner & Sternberg (1985), practical intelligence seeks to fill the gap left by academic intelligence tests in predicting relevant real-life criteria in other than academic settings. For example, Wagner (2000) reported higher predictive validity coefficients of academic intelligence when school-like criterion measures were applied, implicating lower validity coefficients for non-school-like settings.

Central to practical intelligence is the concept of tacit knowledge; it cannot be taught explicitly or sometimes can not be verbalized. According to Henry et al. (2005), tacit knowledge is what one needs to know to work effectively in an environment and involves knowledge about managing oneself, managing others, and managing tasks. *Tacit knowledge about managing oneself* includes knowledge about self-motivation and self-organization. *Tacit knowledge about managing others* refers to knowledge about how to manage one's interpersonal relationships or to function effectively in social interactions. *Tacit knowledge about managing tasks* includes knowledge about how to carry out specific tasks in terms of planning, monitoring, and evaluating one's activities.

In most research approaches, practical intelligence is reduced to the concept of tacit knowledge. However, knowledge, whether implicit or explicit, requires a context-specific conceptualization and operationalization (see Chapter 4.3.2). An adequate measure needs to account for this context-specificity by constructing different tests for different types of occupations, roles, or settings. The problem of context-specificity carries substantial criticisms for the construct (Gottfredson, 2003). Performance depends on personal experience and the opportunity to have acquired the specific knowledge of the respective domain. Gottfredson argues that it is just the strength of academic intelligence to avoid this specificity and, thus, the dependency on culture and experience (see also Süß, 2001; Weber & Westmeyer, 2001). Further criticism concerns the restricted scope of practical intelligence equaling only tacit knowledge. Gottfredson (2003) argued that tacit knowledge has become the new "g", valid in diverse domains and broader than academic intelligence in terms of practical relevance and conceptual scope (see also the next chapter for Sternberg's concept of wisdom as tacit knowledge). Austin and Saklofske (2005) remarked that no "general-purpose practical intelligence test is currently available" (p. 121). Consequently, it is highly questionable whether the conceptualization of practical intelligence as tacit knowledge can at all serve as a new intelligence construct since some of the central requirements for an intelligence construct are violated. Practical intelligence is not generally valid across heterogeneous situations and not only based on cognitive requirements. However, without such a widespread entitlement, tacit knowledge represents an interesting approach to assess domain-specific knowledge. It may be very useful for researchers to attempt to specify social knowledge in specific real-life contexts by the use of the tacit knowledge test paradigm (see Appendix A for an example tacit knowledge test).

4.4.3 Wisdom

"To act wisely in human relations" was one part of Thorndike's definition of social intelligence (1920, p. 218). The concept of wisdom, however, has a long philosophical tradition and was originally located in research on life-span development in the psychological literature. Wisdom was introduced not long ago as an individual differences construct based on a psychometric foundation (Baltes & Smith, 1990; Sternberg, 1998).

The Berlin Wisdom Paradigm

Baltes and colleagues were interested in defining wisdom in terms of the underlying central elements and integrating wisdom in existing individual differences constructs such as academic, social, or practical intelligence and creativity (Kunzmann & Baltes, 2005). Baltes and his colleagues introduced the Berlin Wisdom Paradigm (Baltes & Smith, 1990; Kunzmann & Baltes, 2005) which was derived from philosophical accounts and from implicit theories of laypersons about the concept of wisdom (Holliday & Chandler, 1986; Sternberg, 1985). The Paradigm was intended to offer operational definitions to develop a measure of wisdom-related knowledge. Wisdom was defined as "highly valued and outstanding expertise in dealing with fundamental (i.e., existential) problems related to the meaning and conduct of life" (Kunzmann & Baltes, 2005, p. 117). The range of applications excludes less severe life problems which can be handled by more specific abilities such as social, emotional, or practical intelligence. Moreover, Kunzmann and Baltes (2005) pointed to the integrative character of wisdom which combines cognitive, motivational, emotional, and social components. Five operational criteria for wisdom-related knowledge were formulated that combine intellectual capacities and character (i.e., virtue).

Basic Criteria (inherent to all types of expertise)

1. *Rich factual knowledge* (about human nature and life course)

2. *Rich procedural knowledge* (about ways of dealing with life problems)

Meta Criteria (unique to wisdom)

3. *Lifespan contextualism* (i.e., the awareness and understanding of the many contexts of life, how they relate to each other and change over lifespan)

4. *Value relativism / tolerance* (i.e., acknowledgement of individual, social, and cultural differences in values and life priorities)

5. *Awareness / Management of Uncertainty* (including the limits of one's own knowledge)

Staudinger, Smith, and Baltes (1994) developed a measure of wisdom-related knowledge that required individuals to give verbal responses to prototypical fundamental problems of life (see Chapter 5.3 for a detailed description of the instrument). Several studies were conducted which supported some of the basic assumptions of the Berlin Wisdom Paradigm. People nominated as wise and clinical psychologists showed better performance than age-equivalent groups. More than age, general experiences, professional training and practice, and motivational preferences (Kunzmann & Baltes, 2005) seem to determine performance.

The Balance Theory of Wisdom

As a competing theory, Sternberg (1998) formulated the balance theory of wisdom. Again and analogous to his definition of practical intelligence, Sternberg conceived tacit knowledge as the core component of wisdom. He defined wisdom as a part of practical intelligence which applies tacit knowledge to maximize a balance of others' interests, the context's demands and one's own interests (i.e., the maximization of a common good). In other words, it represents the moral part of practical intelligence (Sternberg, 1998). Furthermore, wisdom is not being taught explicitly and it is "at least partially domain-specific" (p. 356) because of its equalization with tacit knowledge. The value of tacit knowledge as an ability concept central to a new intelligence construct was already discussed in the last section and does not need to be repeated at this point. Hedlund and Sternberg (2000) postulated that research on social, emotional, and practical intelligence and on wisdom can be integrated in the framework of tacit knowledge. Knowledge about managing oneself represents emotional intelligence, and knowledge about managing others is equivalent to social intelligence. Repeating the aforementioned criticisms, it is not adequate to reduce intelligence concepts to the domain of (tacit) knowledge. Therewith, large parts of conceptually and empirically supported differentiations are ignored.

Social Intelligence and Wisdom – Construct Overlap

The construct overlap of social intelligence with wisdom can be addressed in terms of the spectrum of subsumed abilities and in terms of the range of contents and situations in which the construct is applied. With respect to the content domain, Kunzmann and Baltes (2005) conceived social intelligence as more specific than wisdom since it is relevant in more delimited everyday problems. Moreover, the introduction of the dimensions of time and space in the criteria of life contextualism and tolerance represent unique requirements excluded in more specific intelligence constructs. For example, social intelligence is supposed to be relevant in a defined conflict with another person whereas wisdom is responsible for coordinating behavior for solutions that account for a broader viewpoint. However, the opposite position is also justifiable. Wisdom can be seen as the more specific ability because it is restricted to fundamental, serious life problems which rarely occur compared to everyday problems in work and private life and require social intellectual functions. With respect to the spectrum of abilities, wisdom seems to be the broader construct incorporating knowledge, abilities, and meta-cognitive components. Particularly, according to Kunzmann and Baltes (2003, 2005), wisdom-related performance is determined by the interplay and integration of a single intellectual, emotional, or social capacity to deal with a given life problem.

Kunzmann and Baltes (2003, 2005) also elaborated upon the role of emotions for wisdom-related thoughts and judgments and, thus, implicitly addressed the overlap with emotional intelligence. On the one hand, they see emotional reactivity as a meaningful opponent of wisdom-related cognition. Most outstanding life problems also evoke strong emotions that could hinder logical reasoning. On the other hand, they claim that "the ability to work with emotions, to understand emotions, modify them, and use the information they provide to deal with the environment should enhance wisdom-related knowledge in its acquisition and performance" (Kunzmann & Baltes, 2003, p. 337). This idea is similar to the definitions of Branch II (i.e., Facilitation of Thought to enhance intellectual functioning) and IV (i.e., Regulation of Emotion to promote personal growth) of the Four-Branch-Model of Emotional Intelligence. Therewith, the role of Branch II and IV as a type of meta-cognitive ability is stressed (see Table 4.4).

5 Assessing Social Intelligence and Related Constructs

Thorndike's requirement for the assessment of social intelligence was "a genuine situation with real persons" (1920; p. 231). Ironically, ensuing research has tried various approaches to assess social intelligence. Among them, approaches that included genuine situations with real persons were least represented. Social intelligence was assessed by cognitive ability tests predominantly based on verbal material, by the use of self-report inventories of social (cognitive) skills, by implicit theories of laypersons, and sometimes by behavior-based ratings. Throughout its long history, researchers could not agree on a golden rule concerning how best to assess social intelligence, which is reflected within the academic intelligence research. The application of different methods resulted in contradicting validity evidence. This confusion may be responsible for the waxing and waning of research that occurred until today (Matthews et al., 2002).

The present chapter will first present an overview of measurement approaches and the related validity results. First, self-report inventories, implicit-theory approaches, and behavior-based approaches will be presented. Afterwards, the emphasis is put on cognitive ability tests in terms of methodological challenges and problems. The last part of the chapter will present a database of cognitive ability tests of social intelligence and related constructs. The database is extended in the appendix which includes scale descriptions and examples. Some general conclusions about the implications of the various methodological approaches on the validity of the tests will be discussed.

5.1 Overview of Measurement Approaches and Validity Evidence

The following measurement approaches all occurred under the label of social intelligence. However, they were not intended or not capable to assess a cognitive ability construct. Instead, the measurement constructs were social (cognitive) skills and effective social behavior. For completing the picture of social intelligence, they will be reviewed hereafter with a narrow focus on the validity results.

5.1.1 Implicit Theory Approach

The implicit theory approach examines laypersons' conceptions about a psychological construct. The common procedure asks people to identify and describe the behavior of persons who are, according to their implicit theory, high scorers in the respective construct (e.g., highly intelligent). Sternberg et al. (1981) factor-analyzed the descriptions of intelligent people and found three factors labeled *practical problem solving abilities*, *verbal abilities*, and *social competence*. Behaviors classified to the social competence factor included, *accepts others for what they are, admits mistakes, is on time for appointments, thinks before speaking and doing, is sensitive for other people's needs and desires*, etc. Kosmitzki and John (1993) applied the implicit theory approach only to the concept of social intelligence. They extracted three factors labeled *social intelligence*, *social influence*, and *social memory*. Social intelligence consists of the following components: *understanding people*, *social insight*, *perspective taking ability*, *knowing social rules and norms*, *good at dealing with people*, *being warm and caring*, *open to new experiences and ideas*, *social adaptability*, and *being compromising and fair*. Social influence includes *motivation and leadership*, *influence on others*, *dominance and activity*, *manipulating others*. Social memory consists of *memory for names and faces*. Amelang, Schwarz, and Wegemund (1989) used the implicit theory approach to develop a self-report inventory on socially intelligent behavior. They extended the aforementioned approach by having a second sample rate the original behaviors on the dimension of prototypicality for socially intelligent behavior. The result of their studies was a questionnaire of social behavior applicable to self- and peer-report. In a validation study, self- and peer-rated behavior correlated with a general factor of social intelligence which also consisted of self- and peer-ratings.

The benefit of applying this approach on a new intelligence construct can be seen in the investigation of taxonomic foundations when no theory-based accounts are available for this domain. Moreover, the relevance of the construct can be supported by empirical data.

5.1.2 Self-Report Inventories

The trait – ability distinction is most fully developed for emotional intelligence (Austin & Saklsofske, 2005). Social intelligence was repeatedly operationalized by the use of self-report data without such explicit differentiation between ability and mixed models (Barnes & Sternberg, 1989; Brown & Anthony, 1990; Ford & Tisak, 1983; Frederiksen, Carlson, & Ward, 1984; Marlowe, 1986; Riggio, 1986; Riggio, Messamer, & Throckmorton, 1991).

Riggio (1986) constructed the Social Skills Inventory (SSI; Riggio, 1989) which contained six subfacets that resulted from a cross-classification of contents (i.e., social and emotional contents) and skills (i.e., sensitivity, expressivity, and control). The SSI subfacets correlated substantially with personality traits (e.g., social expressivity: outgoing, happy-go-lucky, venturesome, group dependent; social sensitivity: affected by feelings, shy, astute, apprehensive, conservative, tense, undisciplined). High scorers reported more socially effective behavior and richer social contacts. In the study of Riggio et al. (1991), the subscales of the SSI did not show convergent validity with a performance test of social intelligence. Other studies using self-reported social skills aside from social intelligence performance tests succeeded better in showing convergent validity evidence (Barnes & Sternberg, 1989; Ford & Tisak, 1983). Marlowe (1986) identified five dimensions of social intelligence (i.e., prosocial attitudes, social skills, empathy skills, emotionality, and social anxiety) by applying several self-report inventories. These five dimensions did not correlate with academic intelligence assessed by performance data. In a multitrait-multimethod study, Brown and Anthony (1990) found a clear factor structure determined by the applied measurement procedure (i.e., self- and peer ratings) across constructs (i.e., social skills, personality traits, and academic intelligence).

The aforementioned results clearly demonstrate why the use of self-report data is disputable. Empirical evidence has shown that self-report data are not capable to assess a cognitive ability (Paulhus et al., 1998; Riggio et al., 1991). It seems also questionable whether they can serve as a validating instrument without controlling for method-related variance. Nearly no study could convincingly provide evidence for the convergent construct validity of performance measures with self-report inventories. Those that could applied behavior based measures of social abilities.

5.1.3 Behavioral Observations

Behavior-based assessment meets Thorndike's original requirement of genuineness more than any other measurement approach. Behavior-based assessment conventionally asks individuals to behave according to a certain goal in a given setting. Typically, the construct of interest (e.g., social intelligence, social competence, leading competencies, etc.) is operationalized by a-priori constructed rating dimensions. These describe behavioral acts that conform to the supposed manifestations of the construct. Finally, in a test situation, behavior is rated by trained observers according to these behavioral acts.

Ford and Tisak (1983) assessed social behavior in an interview setting as an indicator of social intelligence. Behavior was rated on the following dimensions: the ability to speak effectively, to be appropriately responsive to the interviewer's questions, and to display appropriate nonverbal behaviors. They could prove convergent and divergent construct validity (i.e., with self- and peer-reported social intelligence, and with academic intelligence, respectively). In a study of Frederikson et al. (1984), participants had to take the role of a doctor interviewing his patient. Social intelligence was operationalized by the following dimensions: introduction (i.e., greets client, attempts to put client at ease, etc.), seeking information (i.e., asks about major problems, emotional problems, etc.), giving information and advice (i.e., invites client to ask questions, explains reasons for management decisions, etc.), and affect and support (i.e., expresses understanding, assures client, etc.). Social intelligence as operationalized by the rated behavior correlated only marginally with academic intelligence, proving divergent construct validity.

Stricker and Rock (1990) developed the Interpersonal Competence Inventory (ICI). The ICI is based on a video presentation of an interview between a subordinate and his superior. Participants have to respond orally in place of the superior (i.e., Replies Section). Answers are judged in terms of effectiveness and originality. Also, participants have to write down their description of the situation (i.e., Judgment section). Here, the performance criterion is accuracy. The Judgment Section cannot be conceived as a behavioral indicator whereas the Replies Section operationalizes socially intelligent behavior. Results from correlational and multidimensional scaling analyses showed no coherent evidence of convergent and divergent construct validity.

In social intelligence research, the behavior based approach was rarely applied. Test development and assessment procedures are highly time and resource consuming. Moreover, this approach suits the assessment of social competence and, therefore, of relevant external criteria that validate social intelligence tests. Problematically, every class of situation needs separate rating dimensions for the intended constructs, raters need sophisticated training to detect and rate the desired dimensions, and group testing is not possible. The aforementioned approaches were restricted to interview settings so that the results cannot be generalized. All situations were artificial and the question remains whether people would behave similarly in a genuine situation. Broader approaches to assess social behavior in artificially produced situations represent exercises in the context of assessment center procedures(e.g. role playing, group discussions, etc.). A description of these approaches goes beyond the scope of the present work. For a detailed overview of the typical assessment center structure, exercises, and its validity, see Fisseni and Preusser (2007) and Kleinmann (1997).

5.2 The Cognitive Ability Test – A Methodological Challenge

When the first cognitive ability tests of social intelligence were developed, tests of academic intelligence had recently been introduced. With his claim of genuineness of persons and situations in assessment procedures, Thorndike (1920) stressed the diverse nature of the new construct. The first tests of social intelligence could not match this requirement and strongly compared to those constructed to assess academic intelligence (i.e., Thorndike & Stein, 1937; Moss et al., 1955; O'Sullivan et al., 1965). Item material mainly consisted of artificially produced, decontextualized, mostly verbal descriptions of situations. Obviously, genuineness or social significance requires realistic material that conforms with socially relevant situations (see also Probst, 1973). For example, social interactions or communications are mainly based on the use of spoken and body language as already described in Chapter 4.3.3.3 (see the studies of Mehrabian & Ferris, 1967). The relevance of different communication channels in social situations was not reflected in early ability tests although already Thorndike and Stein (1937) doubted "whether any test which is predominantly verbal can measure social ability" (p. 284). O'Sullivan (1983) focused on the methodological implications of the use of only written language material and expected an unwanted amount of verbal ability variance in tests. Consequently, tests showed substantial correlations especially with verbal tests of academic intelligence. Thus, the autonomy of social intelligence was doubted without acknowledging the possible method-related effect. With today's knowledge about past experiences, with more profound methodological foundations and new technologies, a more deliberate approach to develop assessment procedures of social intelligence is possible which also can account for the genuineness of item material.

Compared to the typical item construction of academic intelligence tests, no comparable body of rules is present or easily constructible for items of social intelligence tests. Neither the item universe is well documented, nor is item material directly available (e.g., compared to number series or word analogies). Social intelligence is usually not explicitly taught in institutional settings. Therefore, a-priori considerations to item construction are helpful and necessary for the selection of item contents, for sampling item material, and for the decision about the formal attributes of items such as item and response format and scoring. The following considerations serve two purposes. First, existing tests of social intelligence shall be described in terms of the aforementioned properties. Second, they shall serve as foundations for the test development of the present work.

5.2.1 Item Origin

The item origin addresses the question of how items (i.e., item material) are sampled or constructed. Tagiuri (1969) distinguished between items produced or sampled in natural or in laboratory settings. More specifically, item material can be distinguished according to the source of the material. It can be provided by real persons, posed by an actor or produced artificially. In a large number of tests of social intelligence, items were produced artificially and only some stemmed from actors and even less from real persons. Artificially produced item material of emotion expressions may typically be a verbal description of a situation, a painted face or a computer-animated face or body presentation. Artificial and posed item material is restricted to be sampled from laboratory settings. Item material from genuine persons can be extracted from natural settings. Tagiuri does not explicate whether natural settings are restricted to direct face-to-face contact between the judge and the target (Tagiuri, 1969) which would make group testing impossible. The present work refers to natural settings as the situation of item material sampling and not of data assessment. With today's technical developments, it is possible to display an earlier recorded close-to-natural setting in the final testing situation.

The use of actors or real persons both bears some advantages and disadvantages. When actors are used, the process of item material sampling can be better controlled in terms of costs and efforts, as well as the match with taxonomic demands. Real material has to be edited in order to make it suitable as an item. Sampling genuine material may not necessarily include the desired taxonomic demands and requires more efforts to create an adequate item from the original recording. However, the crucial question refers to the relevance of the item material for the intended measurement construct and its validity. An example for the ability to understand emotions will be provided hereafter to illustrate the concern. Decoding an artificially produced or posed emotion expression would rarely occur and definitely not at all be relevant in real-life settings. Obviously, it is more relevant to understand the emotion of a real person within the person's natural environment. Emotion expressions by actors or real persons do not only differ in their relevance, they may also evoke different underlying abilities for the accomplishment of tasks. Actors work with conventional emotion expressions which may never be as realistic and manifold or diverse as an emotion expression recorded in a natural setting with a natural background (even by applying a sophisticated script of the situation). By providing information about the real person *in their natural living space*, genuine material clearly enhances the ecological validity of the test material. Moreover, making use of real persons as targets provides the opportunity to sample target information about the person's mental state in the respective situation.

5.2.2 Item Contents and Contexts

The previously mentioned taxonomic foundations help in guiding the selection and the sampling of item material in terms of content and context information. Tests or items can be classified or constructed according to the queried modality (the task product; e.g., emotions, cognitions, personality, etc.), the contents (e.g., written or spoken language, etc.), the setting (e.g., private or public contexts), and the targets (e.g., real persons vs. actors). The possible taxonomic principles will not be repeated at this point. Nevertheless, two aspects need to be accounted for in more detail because of their methodological implications.

Item Contents

Different item contents are conventionally related to the applied task material (i.e., written and spoken language, pictures, and videos). The present work adheres to the distinction of different task material *for the purpose of balancing method-related variance*. Existing multitrait-multimethod approaches differentiate between verbal and nonverbal material. Probst (1973, 1982) distinguished between test approaches that rely on verbal material, pictures, and videos. A different perspective on the impact of item material is the *acknowledgement of separable ability factors* related to different item contents. For example, some researchers addressed the auditory abilities as meaningful contents in academic and social intelligence research (Carroll, 1993; Guilford, 1981; Stankov, 1994; Stankov & Horn, 1980; see, for example, the Test of Implied Meanings by Sundberg, 1966). The doctoral thesis of Seidel (2007) presents a comprehensive approach to develop auditory ability tasks and to integrate an auditory factor into existing models of academic and social intelligence.

Item Contexts

Archer and Akert (1980) show that enough context information is needed in order to give social stimuli their meaning and to allow high performance in social understanding tasks. For example, the statement "Now, it's someone else's turn to organize the family meeting this year! They can eventually also do some work." can only be understood in all its relevance and meaning when people know that the speaker had organized all of the family meetings throughout the last few years and just recently had a heart attack so that he had to refuse from organizing. Thus far, social intelligence testing is criticized for decontextualizing the person (Cantor & Harlowe, 1994; Ford, 1994). It was claimed impossible to be able to validly assess social intelligence without accounting for the task context (see also Ford, 1994). To illustrate the effect of different amounts of context information, Figure 5.1 displays a simple cross-classification of stimuli properties (univocal vs. equivocal) and context availability (none vs. some unspecified amount of context information).

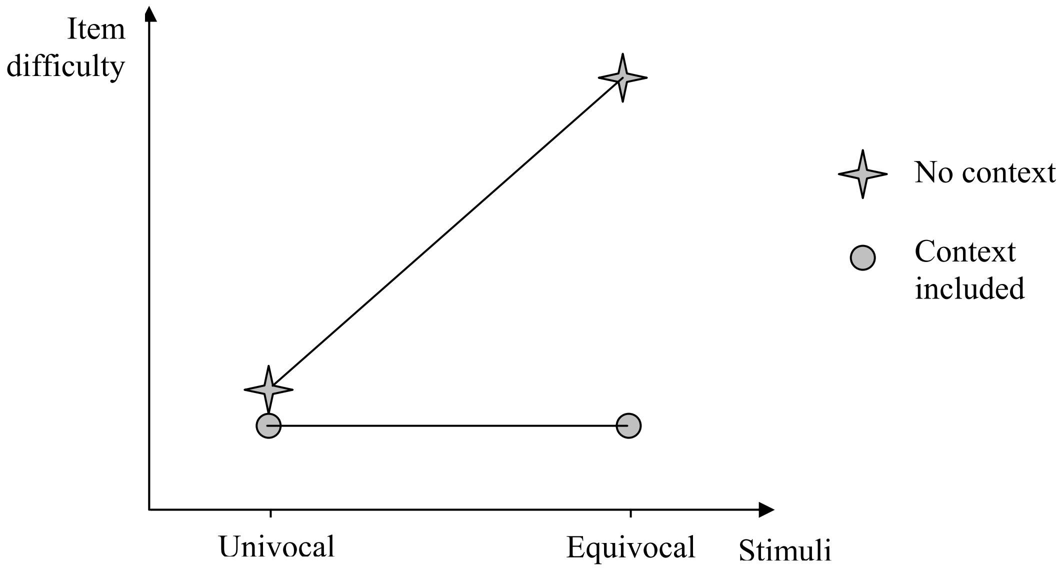


Figure 5.1

Relationship of Stimuli Unambiguousness and Context Information on Item Difficulty

Univocal stimuli allow for the accomplishment of a task without any need to ascertain more information. In this case, item difficulties are low for tasks with or without any context information (e.g., the statement “I am happy” allows for conclusions about the person’s emotional life; a facial expression that can be decoded according to the Facial Action Coding System of Ekman, Friesen, & Hager (2002)may also count to this category). However, most of the relevant social stimuli are equivocal and do not yield an interpretation without additional information (Probst, 1982). For example, a smile may be interpreted as happiness when it is known that the person just passed an exam or a smile can as well be interpreted as irony when knowing that a person is not satisfied with circumstances which he or she cannot change. This represents a simplified illustration of the relationship of context and item unambiguousness to item difficulty.

Context information may not only be a necessary complement for the simple stimulus material. Cantor and Harlowe (1994) claimed that intelligence testing “will miss the within-person, domain-sensitive flexibility and attunement observed in behavioral-observation studies” (p. 160). At the same time, they acknowledged that intelligence can be described as an average capacity which conforms to Carroll’s claim of general validity across situations. To solve this contradiction, balancing the contexts of different items and tasks along taxonomic considerations leads to a control of context-specific variance. This should result in an enhancement of the generalizability of the tasks and thus enhanced general validity. Furthermore, it allows an investigation into whether people differ in their ability to deal with different contexts (e.g., high social intelligence in private contexts versus low social intelligence in work-related contexts). Unfortunately, existing tests of social intelligence vary substantially in how much context information they include and how much they balance different contexts across items or tasks. Some tasks are only located in one context domain, some include very heterogeneous contexts across single tasks without a deliberate balance, while some tasks do not include any context information at all. Thus, in the subsequent presentation of tests, four classificatory elements will be used that concern *(a)* tests without context information, *(b)* tests with only one type of context, *(c)* tests relying on heterogeneous contexts without taxonomic foundations, and *(d)* tests based on heterogeneous contexts and taxonomic foundations.

5.2.3 Item and Response Formats

Researchers were concerned with the overlap of the item properties of social and academic intelligence tasks in terms of their content related and formal characteristics (Neisser, 1976; Schneider et al., 1996; Wagner & Sternberg, 1985). These characteristics are, for example: *the encounter with inconsistent or novel social stimuli, stimuli unrelated to everyday experience, a highly structured task presentation, existence of only one correct answer, and a predefined way to achieve the correct answer.* The content-related aspects have already been discussed in preceding Chapters. The following passages will be concerned with response formats in social intelligence testing. Possible formats are forced-choice formats (e.g., multiple choice items or Likert-based rating scales) and free response formats (e.g., open-ended format or response latencies). Table 5.1 provides example items for all of these response formats and specifies the most central methodological problems that will be addressed subsequently. Within each type of format, various methodological problems appear (see Table 5.1 last column). The problems concerning multiple-choice, open formats, and response latencies, are key issues in every test construction handbook and will not be addressed at this point. The problems related to Likert-based rating scales, however, are of special interest. Ratings-based scales play a prominent role in the context of emotional intelligence testing (e.g., MSCEIT) but are also applied in prototypical social understanding tasks (e.g., the presence of a certain mental state has to be rated on a 1-5-point scale).

Table 5.1

Example Items of Different Response Formats for Tests of Social Intelligence or Related Constructs

Response format	Example item	Typical problems
Multiple choice	Interpersonal Perception Task – 15 (Costanzo & Archer, 1993): The test taker watches a video scene of two adults having a conversation with a two children and has to indicate who the child of the two adults is.	Guessing rate Construction of good distracters
Likert-based rating scale	MSCEIT (Mayer et al., 2002): Test takers see a picture of a facial expression of a person. He has to indicate on a 1-5-point scale how much a certain emotion is expressed in the face (from not at all present to extremely present). Tacit Knowledge Inventory for Managers (TKIM; Wagner & Sternberg, 1991): Test takers have to rate the quality of different strategies for handling a problem in the day-to-day work of a business manager on 1-7-point scale (from extremely good o extremely bad).	Scoring Response biases Group differences Item dependency (when based on the same stimulus)
Open-ended responses	Level of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990): Test taker read a situation description (e.g., “You and your friend are in the same line of work. There is a prize given annually [...] the winner is announced: your friend. How would you feel? How would your friend feel?”). Test takers have to answer the two questions. Answers are rated according to the conformity with different levels of emotional awareness.	Construction of a scoring key Objectivity / interrater reliability
Response latency	Emotional Inspection Time Tasks (Austin, 2004): Test takers have to decide about the emotion displayed in facial expressions as quickly as possible (e.g., Does the face display a neutral mood or happiness?) (i.e., choice reaction time task).	Distributions of reaction time scores

However, the most severe and complex problem in using Likert-based rating scales is that of the scoring procedure (see next Chapter for a detailed account on scoring). Particularly, challenges arise concerning the response biases of individuals and possibly related group differences. For example, Legree, Psotka, Tremble, and Bourne (2005; see also Legree, 1995) considered the effect on the performance of participants who use only part of the rating scale. Legree (1995) suggested procedures that compensate for this effect (e.g., the building of z-scores for each of the items and for the expert ratings and the subtraction of the two).

Particularly, ratings-based scales are associated with the Situational Judgment Test paradigm (e.g., the TKIM in Table 5.1). Situational Judgment Tests (SJTs) are conceived as a class of measurement approaches (Kyllonen & Lee, 2005; McDaniel & Nguyen, 2001). They are applied for the assessment of various constructs and concepts (e.g., practical, emotional, and social intelligence; work-related skills and competences, etc.). They put a focus on the application of contextualized and heterogeneous stimuli. In accordance with Funke and Schuler (1998), McDaniel and Nguyen (2001) assume that the application of contextualized and heterogeneous stimuli enhances the fidelity of item material according to situations occurring in real life. The requirements covered by these tests are not agreed upon by the main protagonists. While the typical conceptualizations claim to measure effective behavior, Situational Judgment Tests are supposed to measure practical know-how (McDaniel & Nguyen). According to Schmidt and Hunter (1993), Situational Judgment Tests simply measure job knowledge; in contrast, Legree (1995) states that they measure everyday problem solving.

The development of Situational Judgment Tests and a respective scoring key is conventionally based on expert statements about critical situations and adequate behaviors in the situations. When experts can agree upon a right solution to a problem, multiple-choice formats can also be applied. When Likert-based rating scales are utilized, answers are judged in terms of the degree of correctness compared with a mean expert statement (i.e., a distance score). One problem of Situational Judgment Tests concerns the fakability when instructions direct the test takers to indicate the most typical or the most likely response (i.e., typical performance). McDaniel and Nguyen (2001) manipulated the instructions in a between-subjects design. Some subjects had to indicate the most likely behavior in a situation (i.e., typical performance), some were asked to indicate the best behavior in the situation (i.e., maximum performance), and some should fake good (i.e., subjects were asked to provide test answers which provoked a test results that presents themselves as favorably as possible). The faking group and the maximum performance group showed performance about half a standard deviation better than the typical performance group. Freudenthaler and Neubauer (2005, 2007) investigated the effect of different instructions on the validity of a Situational Judgment Test. They found that the typical performance condition showed higher correlations with personality traits and self-report inventories whereas the maximum performance condition showed higher correlations with academic intelligence (see Table 5.3 and Appendix A for a more detailed presentation of their test approach). Another problem occurs when more than one item per stimuli is applied. In this case, items are dependent on each other *(a)* as a function of the same stimulus and *(b)* if the instruction promotes a comparison of the single items (e.g., indicate the best and the worst alternative). When separate ratings of effectiveness for each item are applied, the latter problem is eliminated (McDaniel & Nguyen, 2001). The former problem is addressed by the use of specific scoring techniques (e.g., profile scoring; e.g., Snodgrass, 2001), or by defining the rating categories by the use of concrete (behavioral) anchors.

To recur to the question of response format, MacCann (2006) investigated the effect of different response formats on the between-constructs' correlations of emotional and academic intelligence. Tests to assess these constructs utilized multiple-choice and ratings-based formats. Results showed larger construct overlap when the same response format was applied. Funke and Schuler (1998) investigated the effect of the response format on the criterion validity of a situational judgment test of social competence. Results generally supported an open response format. The authors attributed this to the higher fidelity of the responses in terms of the behavioral requirements of both the construct and the assessed criterion. In general, the applied items and response format influence the construct's convergent and divergent validity. Consequently, early tests of social intelligence which applied traditional response formats (typically multiple-choice) were likely to show an unwantedly large overlap with tests of academic intelligence due to them using the same type of response format.

5.2.4 Scoring

According to Carroll (1993), a cognitive ability task demands the *correct* information processing as a critical condition for successful performance. The scoring key has to provide an objective rule for judging the response to a test item as correct or as more or less correct (i.e., degree of correctness on a continuum; Nevo, 1993; Wilhelm, 2005). Guttmann and Levy (1991) listed three types of rules that allow for the deduction of a correct answer to an intelligence test item: logical rules (i.e., typically mathematical problems), scientific rules (i.e., typically extracted from experimental results, or in other words, physical reality), and semantic rules (i.e., extracted from the dictionary). Relying on the results of an implicit theory approach, Nevo (1993) added the agreement with an authority (i.e., common knowledge or consensus) as another scoring rule to the three rules of Guttman and Levy (1991).

To identify the correct answers to a social intelligence test item, no logical or semantic rule can provide enough information. In contrast to academic intelligence testing, what has to be inferred about can rarely be perceived directly (e.g., physical vs. psychological attributes). Consequently, scientific rules or authority agreements must be applied. It is possible to apply a scientific rule when science has proven the truthfulness of the rule and when it can be applied one-to-one on the respective stimuli. For example, science has established certain rules about the appraisal and the expression of emotions which may be applied on narrowly defined social stimuli (e.g., the Facial Action Coding System, Ekman et al., 2002). A scientific rule is also applied when a correct answer is based on observable facts (e.g., a person who was present in the stimuli is remembered correctly). However, scientific rules are no longer valid when a certain range of interpretation is possible or, when not all of the necessary information is available (e.g., anger follows frustration only under certain conditions related to the involved persons and the situation) (see also Mayer et al., 2001).

Initially, the agreement with a person of authority is a vaguely composed rule that depends on the conception of the authority figure. In this context, Nevo (1993) generally speaks of consensus or common knowledge as criteria. Cline (1964) applied a more fine-grained distinction between three sources (i.e., authorities) of criterion information to identify the correct answer to a test item: experts, associates (i.e., peers, subordinates, spouse, teachers, bosses, acquaintances, etc.), and the self. Comparably, Kenny (1994) differentiated between several sources for the criterion information: self-report, group consensus, experts, behavioral observations, and operational criteria. He referred to operational criteria as objective social facts such as “the two persons are siblings” (i.e., standards-based scoring or scientific rules). Apart from the target scoring and standards-based scoring procedures, most common approaches are anchored in emotional intelligence research, and distinguish between group and expert consensus scoring (MacCann, Roberts, Matthews, & Zeidner, 2004; Mayer et al., 2002; Tagiuri, 1969). It is assumed that the general consensus between (large) groups of participants or experts is founded on a common knowledge base in the population that reveals the correct answer (Legree, 1995; Mayer, Salovey, Caruso, & Sitarenios, 2003).

The construct definitions of social memory and social perception require the objective presence of the queried stimuli. According to these definitions, the scientific rule can be applied by relying on the objectively present information (e.g., a response to a memory test item is right when the stimulus showed exactly the response that was asked for). Perception tasks are frequently scored in terms of response latency (e.g., a test taker indicates by a keystroke that he or she has detected a target of perception, the keystroke is then scored by the reaction time between the target presentation and reaction). The definition of social creativity requires that scoring accounts for the number and diversity of responses. Thus, correctness does not play a role in this ability domain. The construct definition of social understanding is not as directive. This scoring is far more intricate and it depends upon the nature and complexity of the queried information (see Chapter 4.3.3.2 for an overview). This can be a feeling or a thought of an individual. It can also be a more general concept such as a personality trait, future behavior, a relationship between two or more people, or a solution or explanation to a complex problem. Existing approaches in the literature refer to the wide spectrum of possible social information and consequently claim that with the rising complexity of the stimulus, the identification of a correct answer gets harder (Matthews et al., 2002). Schulze, Wilhelm, and Kyllonen (2007) even claim that "the unavailability of indisputable rules can be considered as the most important problem in emotional intelligence research" (p. 212).

Table 5.2 presents examples of different queried modalities and scoring options of prototypical social understanding tasks. Obviously, scientific rules are not applicable for most of the examples and depend on the availability of objective information. Neither standards-based nor target scoring can be used when the queried information represents solutions or explanations to social problems. The problem becomes obvious when imagining that the typical answer to the example item of the Chapin Social Insight Tests is "It depends ...." (see also Bless et al., 2004). Consequently, only group and expert consensus scoring is possible. The passages that follow will focus more specifically on the three most prominent and discussed scoring procedures in social intelligence testing: target scoring and group and expert consensus scoring.

Table 5.2

Queried Modality, Examples, and Related Scoring Options of Social Understanding Tasks

Queried modality	Example item	Scoring options
Emotions	MSCEIT (Mayer et al., 2002): Test takers see a picture of a facial expression and have to indicate how much a certain emotion is expressed in the face.	SBS (limited item contents), TS, GCS, ECS
Thoughts	Fictitious: Test takers listen to a phone call (only one side of the conversation done by the target). Several thoughts have to be rated in terms of how much the target would agree with to have had during the phone call (e.g., “The person I'm talking to just has no interest in trying to understand me.”).	TS, GCS, ECS
Relation-ships between others	A. Interpersonal Perception Task – 15 (IPT-15; Costanzo & Archer, 1993): Test takers watch a video scene of two adults having a conversation with a two children and has to indicate which one is the child of the two adults is. B. fictitious: Test takers listen to a conversation between a pair of people and has to judge the relationship between the two with regard to sympathy, familiarity, etc. from the perspective of one of the two	SBS (only Example A., when social facts are available), TS GCS. ECS
Personality traits	Fictitious: A test taker has to rate the personality traits of strangers on the Big-Five-dimensions.	TS, GCS, ECS
Future behavior	Fictitious: A test taker has to identify the most likely future behavior of a target out of several alternatives.	SBS, TS, GCS, ECS
Problem explana-tions or solutions	Chapin Social Insight Test (Chapin, 1967; Gough, 1968; see Appendix A): The test taker reads a situation description: “A man bought an expensive automobile [...] he gave several reasons [for the purchase], but the one reason he did not give was [...] Test takers then have to choose out of four alternatives the one that represents the most logical explanation to the problem.	GCS, ECS

Note. SBS: standards-based scoring relying on scientific rules
TS: target scoring
GCS: group consensus scoring
ECS: expert consensus scoring

5.2.4.1 Target Scoring

The classical test approach to social understanding (i.e., interpersonal perception) relies on the accuracy score (Bronfenbrenner et al., 1958; Cronbach, 1955; Tagiuri, 1969). The accuracy score represents the typical target scoring procedure. Target scoring allocates the difference between a target's and a judge's answer to the score. Conventionally, the squared difference or a reverse difference score is used in order to give a high performance a nominally higher score. This scoring is based on the idea that the target has more information about its own mental states than any outside observer (Mayer & Geher, 1996). Target scoring is restricted to item contents that are concerned with the mental states of the target. Targets can report about their emotions, thoughts, motivations, or intentions. The most crucial criticism of target scoring concerns the bias of the target in judging its own mental state. A feeling may be too complex to be communicated, the target may miss an adequate label for the feeling or thought, or the target may intentionally report a more socially desirable feeling or intention, etc. (MacCann et al., 2004; Mayer & Geher, 1996). O'Sullivan (2007) reported a study by Spain, Eaton, and Funder (2000, as cited in O'Sullivan, 2007) which had participants (in this case equaling the targets) and peers or acquaintances rate *(a)* the participant's daily emotional experiences and *(b)* behavioral indicators related to personality traits. Targets were more accurate than acquaintances in judging their emotional experiences but acquaintances were more accurate in rating the target's behavioral indicators related to the personality traits. O'Sullivan concluded that "self-reports are clearly better for the prediction of emotional experience, while for (overt) behavior the picture is mixed" (p. 264).

Other criticism relates to the dimensionality of the target score. Bronfenbrenner et al. (1958), Buck (1983), Cline (1964), and Gage and Cronbach (1955) doubted that a person's accuracy in judging others can be described in terms of a single dimension or a single component score. Figure 5.2 displays the possible variance sources that contribute to the deviation between the target's and the judges' answers for a prototypical Likert-based rating scale. The target answer in Figure 5.2 is "5" on a 7-point scale. Three possible judges' responses are marked (x=2; y=3; z=6).

As the most evident alternative, the variance of target scores can be explained first and foremost completely by the difference in ability between the different judges providing ratings. The ability of a judge can be determined regardless of item difficulties, stimuli or target properties since a maximum strive for accuracy in typical performance testing should compensate for any of these influences.

Against this, Buck (1983) distinguished between *perception* and *knowing* as processes that contribute to the accuracy of a judgment (also bottom-up vs. top-down processing, respectively; Bless et al., 2004). Comparatively, Bronfenbrenner et al. (1958) and Cline (1964) distinguished between the *sensitivity to the individual* (*differential accuracy*) and the *sensitivity to the generalized other* (*stereotype accuracy*). Both differentiations claim that judges may either rely on the perception of the available stimuli (i.e., what is specific about the target and different compared to what one knows), or on what they know about the target, the target's social group, or the situation. The two ability domains are supposed to be independent. Cline (1964) investigated the relationship of the stimulus information to stereotype accuracy and differential accuracy and reported an increase in differential accuracy with an increase in information. Conversely, stereotype accuracy decreased while the amount of information increased. Thus, to what extent both ability domains explain variance in the target score depends on the availability of relevant information (i.e., few information leads to top-down / knowledge controlled information processing) and the amount of prior knowledge about the target and the situation.

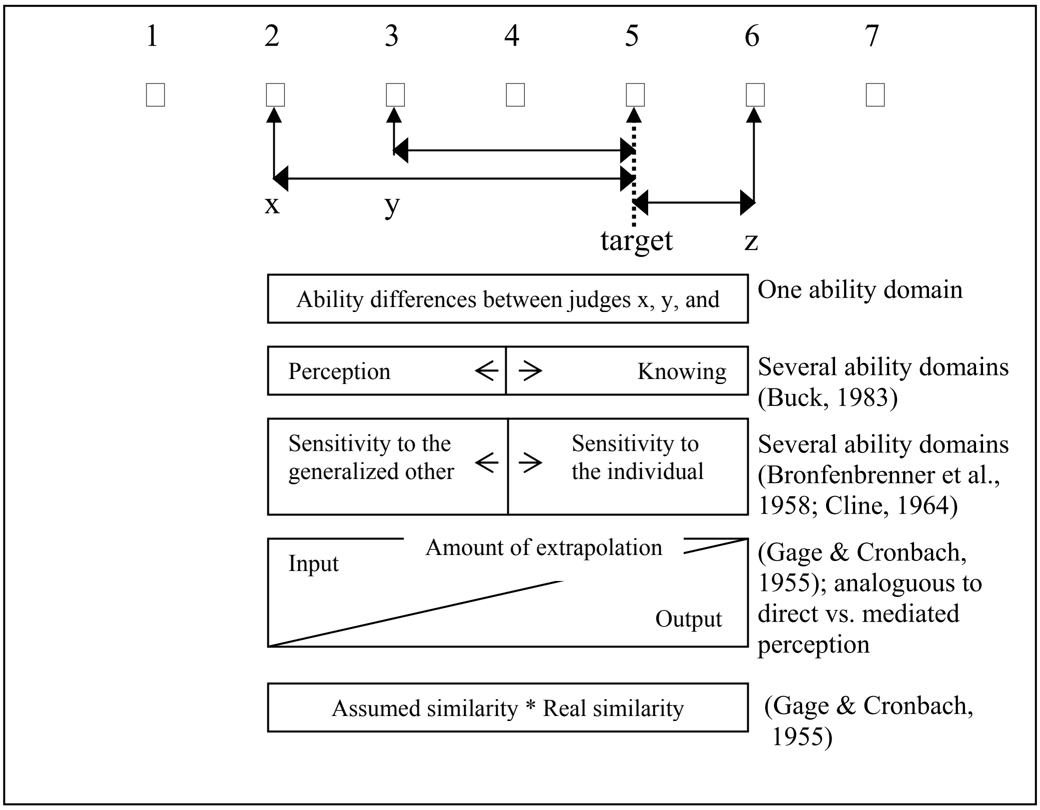


Figure 5.2

Example Rating Scale Indicating the Variance Sources of the Target Score

Gage and Cronbach (1955) were concerned with the degree of extrapolation or inference from the *input* information to the requested *output* of social judgments. Items can put demands either on the input process (perceiving without acquaintance with the target) or on the interpretation process (extrapolation from stimuli to the requested output). Contrary to Bronfenbrennner et al. (1958) and Buck (1983), Gage and Cronbach's ability domains are not independent variance sources and knowledge (i.e., as a form of stereotype accuracy or sensitivity to the generalized others) does not play a role in the accuracy of a judgment. As long as the input directly reflects the requested output, no extrapolation is necessary. Extrapolation from input to output becomes indispensable when not enough information is included in the stimuli. Comparably, Buck (1983) introduced the concepts of direct and mediated perception which also reflect the proportion of input to requested output (see Chapter 4.3.3.1). Gage and Cronbach were further concerned with the influence of the similarity between the judge and the target. They assume high performance when judges assume high similarity (i.e., the responses then reflect the judges' own way of thinking and feeling) while also the real similarity is equally high. Consequently, high performance also results from a combination of low assumed and low real similarity.

Cronbach (1955) was the first to establish a componential analysis system to decompose the target score into different variance sources in order to account for the real and the assumed similarity in the use of rating scales. The system incorporated four elements that add to the calculated judge-target difference. They shall be described briefly without focusing on the mathematical issues: *(a) Elevation* measures the difference between the mean of a judge's responses for all targets on all items and the mean of the targets' answers on these items. *(b) Differential Elevation* measures the extent to which a judge predicts the deviation of the mean of all target answers over all items from the mean of one individual target. *(c) Stereotype Accuracy* measures the degree to which a judge predicts the mean answer of a group of targets. *(d) Differential Accuracy* measures the extent to which the individual target answer deviates from all other targets on one item. A necessary requirement for such an analysis is the application of the same items for every target (e.g., every target is judged on the same personality dimensions). It was acknowledged by researchers that this analysis and thus, the final *Differential Accuracy* score is difficult or impossible to interpret (Cline, 1964; Colvin & Bundick, 2001). Cline (1964) demanded a more theory-driven perspective on the process of judgments that accounts for the interaction of the judged trait, possible constant tendencies of the judge in judging this trait, and the effect of the target being judged on this trait. Cline listed even more variables that may bias any judgment: social desirability, similarity, the tendency to agree with a statement ("Yes"-tendency), the use of stereotypes, personal reactions such as sympathy, use of implicit personality theories (e.g., the assumption of a relationship among separate traits or items judged), and the tendency to make extreme or central ratings. Colvin and Bundick (2001) sampled some design and statistical techniques to deal with the Cronbach components. For example, they recommended the use of forced-choice rating techniques such as the Q-sort technique to constrain the ratings of all judges to the same mean and variance to remove the *elevation* factors. Alternatively, the standardization of ratings is supposed to have the same effect (i.e., or correlations-based scores; e.g., Snodgrass, 2001).

Kenny, West, Malloy, & Albright (2006; see also Kenny & Winquist, 2001) summarized the problems and advantages of both componential and noncomponential approaches. They acknowledged that science has a strong interest in the overall degree of accuracy (noncomponential analyses) which is a direct function of the judge's performance. They saw the application of componential analysis as restricted to specific research designs (i.e., several judges judge several targets on the same items). Moreover, the estimation of the different components results in the addition of error und thus, in a less reliable measurement. In agreement with Cline (1964), Kenny et al. (2006) state that componential analysis may be too complex and not possible to conduct by accounting for every possible component. However, Kenny et al. conclude that whenever appropriate and possible, componentional analysis should be conducted (see also Bernieri, 2001). In contrast, Funder (2001) strongly claimed to prefer design controls instead of possibly biasing statistical controls. According to Funder, statistical controls should not be conducted blindly or inconsiderately and the possible negative consequences should be accounted for (see Chapter 9.2.2 for the discussion of the debatable issues).

Besides the componential analysis as suggested by Cronbach (1955), Cline (1964) or Kenny et al. (2006), Snodgrass (2001) proposed correlations-based scoring as an alternative procedure to deal with ratings-based scales. The correlations-based scoring method is conventionally applied in interaction research and is a subtype of target scoring (Bronfenbrenner et al., 1958; Funder, 2001; Ickes, Stinson, Bissonnette, & Garcia, 1990; Kenny & Winquist, 2001; Snodgrass, 2001). Correlations-based scoring assigns each subject (i.e., the judge) a score based on the correlation between the subject's and the target's answers on a set of items. Therewith, the information from the single item is lost. Interaction paradigms typically involve the same persons as both, targets and judges, being put into an encounter. The scoring method is sought to account for an interaction effect of the accurate sending and perceiving of cues (Ickes et al., 1990; Snodgrass, 2001). Snodgrass (2001) further claims that correlations-based scoring compensates for different rating tendencies of judge and target by relying on z-standardized scores. For example, if a target person tended towards extreme answers and a judge towards the middle of the scale, this would be equaled with inaccuracy in the traditional target scoring procedure. The predominant disadvantage of this scoring method, however, represents the loss of information by single items and thus, the lack of reliability information.

In conclusion, the differential use of ratings-based scales of a judge and a target, the effect of similarity between judge and target, and the question of the validity of the target's answers, need to be considered when target scoring is applied. Research can be designed so that the aforementioned problems may possibly be controlled for: *(a)* The targets should represent no prototypical representatives of a social group. *(b)* Enough background information about the situation and about the person should be provided. *(c)* Stimulus material should provide enough information so that the relevant cues can be perceived by every judge. This should minimize the variance in the target score explained by perceptual abilities and restrict performance to interpretation requirements without top-down controlled components. More design issues will be addressed in the last Chapter (9 and 10) also concerning some research designs and questions relevant for the present problems.

5.2.4.2 Group Consensus Scoring

There are several algorithms to calculate consensus-based scores. MacCann et al. (2004) analyzed the effect of different algorithms on the psychometric quality and validity of emotional intelligence scales. Results suggested an advantage of the so-called proportion or mode scoring procedures. Proportion scoring is typically used in emotional intelligence testing (MSCEIT; Mayer et al., 2002) and allocates a score to a response according to the proportion of people in the sample endorsing that response. Mode scoring assigns correctness to the modal response in a sample while all other responses receive a score of zero.

Emotional intelligence research is confronted with several severe criticisms about group consensus scoring. For example, an analysis of item difficulties is not possible (Legree et al., 2005; Schulze, Wilhelm, & Kyllonen, 2007). When this scoring is applied on a difficult item, a low score would result for a highly able person (Matthews et al., 2005; Schulze et al., 2007). If a more able person accomplishes a hard item while the person with lower ability does not, the less able person receives a higher score (.80 vs. .20) because a hard item is conventionally solved correctly by a minority of persons. Thus, scoring depends on the mean level of ability in the sample. Legree et al. (2005) demonstrated different distributions of Likert-based and multiple-choice items scored consensually with respect to different levels of expertise. They could show that the distribution parameters of central tendencies (typically higher for a higher level of expertise) and variance (typically lower for a higher level of expertise) vary with expertise differences in samples. Thus, one of many problems arises if the level of expertise is unknown and not equally distributed in an applied sample. Tagiuri (1969) concluded that group consensus scoring favors the judge who agrees with the average response of the comparison group but, perhaps agrees with the most common biases found among a sample (i.e., stereotypes or judgment biases).

Finally, Kyllonen and Lee (2005) claimed that group consensus scoring should only be applied when unusual judgments are required for which no expert statements or standards-based rules are available (e.g., judging the emotions in pieces of arts or in music). The aforementioned critical points dominate the discussion about emotional intelligence and raise doubts that this type of scoring is capable to objectively assess an intelligence (see also Matthews et al., 2002; Roberts et al., 2001).

5.2.4.3 Expert Consensus Scoring

An expert is supposed to be a specialist who understands the internal states of an individual better than the individual him- or herself and better than any outside observer. Expert consensus scoring is sometimes applied in social intelligence research (e.g., the Four Factor Test of Social Intelligence, O'Sullivan & Guilford, 1976) and is now prominent in existing approaches to assess emotional intelligence (e.g., MSCEIT, Mayer et al., 2002; Situational Judgment Tests). Typically, experts are nominated by test developers. They respond to the items with their assumed privileged knowledge base. This response becomes the criterion against which the participants' answers are scored as correct (in terms of an absolute value or in terms of the deviation from the experts' opinion).

The role of expert consensus scoring is also subject to critical debate. For instance, Schaie (2001) and Legree (1995) require the application of a large panel of experts and a satisfactory degree of agreement between the experts (conventionally 75%). For example, the size of expert samples two (for the Multifactor Emotional Intelligence Scale, MEIS) and 21 (for the MSCEIT). The resulting correlation between expert and group consensus scoring varies from moderate correlations for only two experts and a close-to-perfect correlation for 21 experts (r = .26 and .96, respectively) (Roberts et al., 2001). Possibly, this result is just a function of the size of the expert sample and suggests that expert ratings converge with the group consensus when enough experts are tested. However, in some circumstances, an expert also may provide just a more reliable indicator of a group consensus (Legree, 1995; Mayer et al., 2001). Mayer & Geher (1996) also considered circumstances in which an expert would provide a more accurate judgment than the group because of a richer knowledge base.

Integration of Scoring Procedures

In summary, every scoring procedure bears advantages and disadvantages. The scoring key for group consensus scoring is quite easily and economically achieved compared to expert consensus and target scoring. Group and expert consensus scoring allow for the application of a broader range of item material or contents beyond the judgment of a target's mental state (see Table 5.2). But, economy of efforts cannot be the crucial point to decide about the adequacy of a scoring procedure. With respect to the standards of intelligence testing, target scoring seems to provide the most objective scoring procedure when standards-based scoring is not possible. Expert consensus scoring also seems promising when item contents are applied that cannot be judged by the use of target scoring (e.g., complex social situations when no targets are available). Moreover, expert ratings can be used as a validation criterion for the target responses. Colvin and Bundick (2001) even demanded the application of multiple criteria for judging accuracy.

However, of particular interest, is the concern that the validity of a construct is dependent upon on which scoring method is utilized. What makes a person score high or low depending on the selected scoring method? Mayer and Geher (1996) summarized research results suggesting a nonsignificant correlation between target and group consensus scoring. They reported empirical evidence against a target-group consensus convergence based on simple emotional judgments (mean $r = .30$). The authors even expect a drop in correlation size for more complex social cues. Davies et al. (1998) applied several tests of emotion perception based on different item material and applied target and group consensus scoring. They reported lower reliability coefficients for group consensus scoring. The two scoring procedures correlated substantially with $r = .99$ for a test of emotion perception based on pictures and $r = .48$ for a test based on auditory voice presentation. Interestingly, other emotion perception tests that did not allow target scoring (e.g., emotions expressed in colors or musical excerpts) showed unacceptable reliability coefficients ($r = .28$ and .37, respectively) of group consensus scoring suggesting that there was no commonly agreed-upon criterion (i.e., no common knowledge base) in the applied sample. No investigations are available in the literature that report about the agreement between target and expert consensus scoring.

The upcoming passages are intended to elaborate the relationship between target and group consensus scoring depending on item difficulty for a virtual sample of 20 subjects. The same type of analysis should be repeated later on based on data from the empirical studies. At present, it should be exemplified by four items (all belonging to one scale) how the relationship of group consensus scores to scales scored according to any external standard (such as a target or experts' answers) appears for items of different difficulty. Therefore, Figures 5.3 and 5.4 present the effect of item difficulties on the bivariate distributions of scales scored by target and group consensus scoring. The response format is a 7-point rating scale. Target scores were calculated by the negative difference between the target's and the judge's answer, and consensus scores by proportion scoring. Figure 5.3 displays four different items, the target responses were "6", "5", "5", and "1", for Items 1, 2, 3, and 4, respectively. Items 1 and 2 are items of low difficulty (*m* = -.600), the mode response of the sample (which receives the highest score in consensus scoring) converges with the target score (receiving the highest score in target scoring). In Item 3 and 4, the difficulties are larger (*m* = -2.000 / -1.500, respectively). The mode response of the sample does not converge with the target answer. Figure 5.3 also presents the means, standard deviations, and item-total correlations for all items. In the graphs on the left side of Figure 5.3, for each item, the frequencies of the raw responses for each rating category are displayed. The proportion per bar (i.e., the relative frequency of this category in the sample) equals the group consensus score in the graphs on the right side. There, the bivariate scatterplot of each item based on target (x-axis) and group consensus scoring (y-axis) is presented.

A linear relationship between target and consensus scores on the item level only emerges when the mode response converges with the target response (Item 1 and 2). In Item 1, a target score of "-1" (i.e., for one point difference from the target answer) is assigned to the responses "7" and "5". However, scoring Item 1 with group consensus scoring results in different consensus scores (i.e., "0.05" for a response of 7; "0.2" for a response of 5) although the same target score is assigned. This happens because the categories are represented by a different relative proportion of responses in the sample. When the mode response does not converge with the target answer (Item 3 and 4), a curvilinear relationship in the form of a reversed "U" between target and consensus scores is discovered, both for normal and for skewed distributions of the raw scores.

Item 1
target scoring
m = -.600
sd = .883
r_{it} = .830
group consensus scoring
m = .415
sd = .236
r_{it} = .581

Item 2
target scoring
m = -.600
sd = .681
r_{it} = .907
group consensus scoring
m = .335
sd = .175
r_{it} = .536

Item 3
target scoring
m = -2.000
sd = -1.124
r_{it} = .897
group consensus scoring
m = .260
sd = .123
r_{it} = .443

Item 4
target scoring
m = -1.500
sd = .889
r_{it} = .891
group consensus scoring
m = .325
sd = .133
r_{it} = -.002

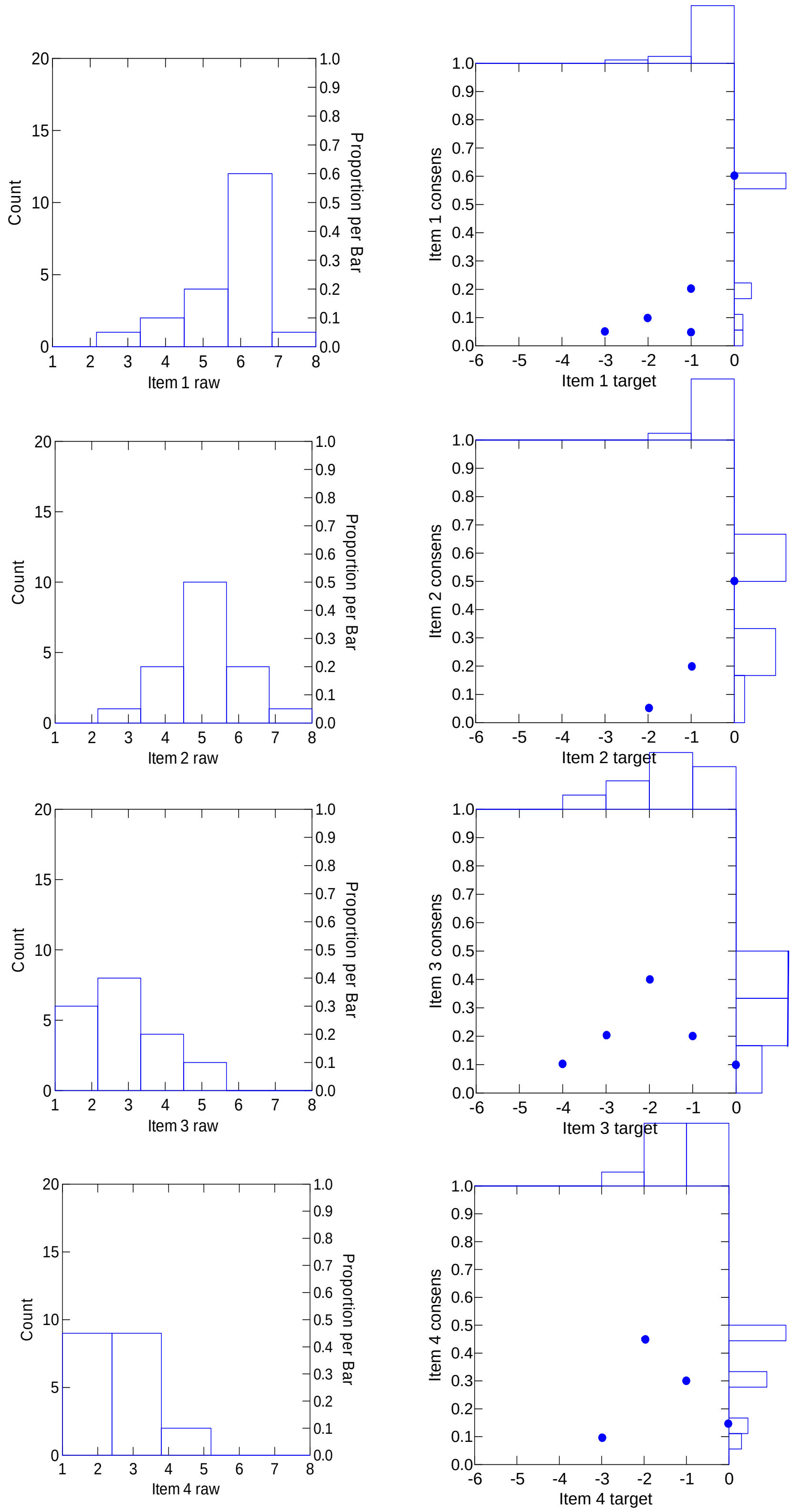


Figure 5.3

Distributions and Bivariate Scatterplot of Four Exemplary Items (Based on Target and Group Consensus Scoring)

Note. Target responses were “6”, “5”, “5”, and “1” respectively for Item 1 - 4

Aggregating Items 1 to 4 results in the bivariate scatter plot displayed in Figure 5.4. The reliabilities of these scales is Cronbach's alpha = .940 (target scoring) / .579 (group consensus scoring). The correlation between the target and the group consensus scores is r = .785. However, this number does not seem to be a good indicator of the true relationship looking at the curvilinear relationship.

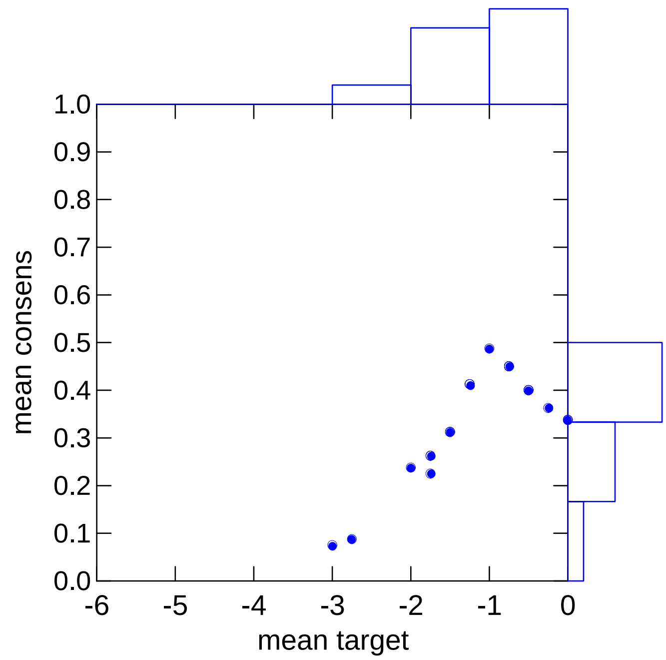


Figure 5.4

Bivariate Scatter Plot of the Relationship of Target and Group Consensus Scoring of One Scale

Importantly, the final group consensus score is a function of the underlying distribution of the raw score in the sample. Thus, high scores in group consensus scoring are not necessarily associated with a high score in scales scored by an external standard. This illustration could show that, in this virtual sample, the bivariate distribution of target and group consensus scores depends on the item difficulty which, in turn, depends on the sample. In correspondence with the literature, this demonstrates a lack of objectivity for the group consensus scoring procedure. Please also note the corresponding analysis presented in Chapter 8.4.4.3 based on the empirical sample of Study 2.

5.3 A Database of Cognitive Ability Tests

The present chapter presents existing test approaches that assess social intelligence and related constructs. The tests are presented in Table 5.3 in chronological order. For economic reasons, tests were excluded that represent predecessors of current instruments (e.g., the MEIS as a predecessor to the MSCEIT is not included). Table 5.3 summarizes the intended measurement constructs, the taxonomic classifications of the test, and the basic psychometric properties of the scales. The taxonomic classifications are derived from the performance model of social intelligence including operative and content-related components and a simple classification regarding context. The database does not lay claim on completeness because not all the literature was available or the literature did not provide the relevant information. Thus, either some entire approaches or single information to one approach are surely missing. Areas where there is missing information connotes that the reviewed literature does not provide any information. In some instances, this means that a classification could not be done. The reliability and validity coefficients are derived from the manuals if they were available or from empirical studies. For some tests, there may surely be more validity studies available in the literature so that the reader who knows the literature may sometimes miss one or the other result. However, it was attempted to make a representative selection of research results in order to provide an adequate overview.

Table 5.3 is complemented by a database of tests in Appendix A that includes scale descriptions and example items. Two tests are referred to within the subsequent text because they represent comprehensive test batteries that are worth a more thorough look: the MSCEIT (Mayer et al., 2002) and the Four Factor Test of Social Intelligence (O'Sullivan & Guilford, 1976). Please note that not all the information in the Table and the database is interpreted or referred to subsequently. Many of the information was sampled in order to provide a collection of test approaches that allows to be consulted any time after the present work.

Table 5.3

Overview of Tests of Social Intelligence and Related Constructs Including Taxonomic Classifications and Psychometric Properties

Test and subscale	Intended construct	Taxonomic classifications			Scoring[d]	Item format/ No. of Items[e]	Item origin[f]	Reliability	Validity[g]	
		Operation[a]	Content[b]	Context[c]					Convergent	Divergent
Dymond Rating Test of Empathy (Dymond, 1949; O'Sullivan, 1983, Orlik, 1978; Walker & Foley, 1973)	Person perception / Role taking	U	V	HE	TS	R / -	G	$r_{split\ half}$ =.82	-	-
George Washington Social Intelligence Test (GWSIT; Moss et al., 1955; Thorndike & Stein, 1937)	Social Intelligence							r_{retest}=.89 $r_{odd/even}$=.88	mean ***r* = .34** within scales (Judgment and Recogn.) [21 Study 2] *Judgment*: ***r* = .20 / .04 / .12**with ST/EG/CP[21 Study 1] *Recognition*: ***r* = .06** with EG (4-Factor-Test) [21 Study 1]	Overall: ***r* = .69** w. overall AI mean ***r* = .34** with AI scales (highest with verbal subscales of AI)[21 Study 2]
Judgment in Social Situations		U	V	-		M / -				
Recognition of Mental State Behind Words		U	V	-	-	M / -	-			
Observation of Human Behavior		K	V	-		M / -				
Memory for Names and Faces		M	P	NO		M / -				
Sense of Humor		U / K	V	-		M / -				
Role Taking Test (Feffer, 1959; Orlik, 1978 and Walker & Foley, 1973)	Role taking	U / C	B	NO	Open response ratings	O / -	AR	Interrater: .89	-	-
Test of Implied Meanings (Sundberg, 1966, see also Stricker & Rock, 1990)	Decoding nonverbal communication	U	A	-	-	- / -	-	$r_{split\ half}$ =.62	***r* = .08** with SR SI[19]	***r* = .00 / .39** with numeric / verbal AI[19]
Chapin Social Insight Test (SIT; Chapin, 1967; Gough, 1968; Keating, 1978; Weis & Süß, 2007)	Social insight	U	V	HE	GCS	M / 25	G	Weis and Süß, 2007 Keating (1978): α = .42/.84	***r* = .22** with ST (4-Factor-Test) ***r* = .15** with IPT-15 ***r* = .06** with Faces Test (MSCEIT) [20] ***r* = .09 - .29** with SR SI	***r* = .07 - .40** with AI scales (lower for figural tests)[20] zero correlations with personality
Communication of Affect Receiving Ability (CARAT; Buck, 1976, 1983; Stricker & Rock, 1990)	Nonverbal receiving ability	U	F	NO	SBS / TS	R / M / 32	G	α = .56 r_{retest} = .79 $r_{odd/even}$=.19	Buck, 1983: ***r* = .04** with PONS ***r* = .24** with Video PONS ***r* = .12** with SR SI	***r* = .03 / .10** with numeric / verbal AI[19]

Table 5.3 *continued*

Four Factor Test of Social Intelligence (O'Sullivan & Guilford, 1976; O'Sullivan et al., 1965; see also Barchard, 2003)	Behavioral cognition of...							Barchard / O'Sullivan et al.:	Riggio et al. (1991)[16]: ***r* = .28 - .44** within tasks ***r* = .09 - .19** with SR SI	Riggio et al. (1991)[16]: correlations with verbal / figural AI:
Expression Grouping (EG; CBC4)	Classes	U	P	NO	ECS	M/26	AR	α=.20 / .62	Wong et al. (1995)[21 Study 1]	EG: **.19-.29 / .19**
Missing Cartoons (MC; CBS1)	Systems		P	HE	ECS	M/28		α=.57 / .77	EG: ***r* = .06** with GWSIT subscale)	MC: **.21-.45 /.39**
Social Translation (ST; CBT4)	Transformat.		V	NO	ECS	M/24		α=.71 / .86	ST: ***r* = .28** with EG	ST: **.43-.52 / .30**
Cartoon Prediction (CP; CBI3)	Implications		P	HE	ECS	M/29		α=.46 / .79	CP: ***r* = .30/.25** w. ST/EG)	CP: **.23-.33 / .23**
Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Buck, 1983; Hall, 2001)	Nonverbal sensitivity	U		NO	SBS	M / 220	AC		***r* = .50** with MERT, JACBART[2]	Rosenthal et al. (1979): ***r* = .14** with AI
full length								α = .86	Buck (1983) / Hall (2001): ***r* = .04** / **.16** with CARAT	***r* = .22** with personality
Video PONS			A					α < .40	***r* = .24** with CARAT	(*median r* of various samples)
Audio PONS			F					α < .30	***r* =** / **.20** with IPT-15	
Couples Test (Barnes & Sternberg, 1989)[3]	Nonverbal decoding skills	U	P	NO	SBS	M / 24	G	α = .34	***r* = .14** with supervisor task (same test principle) ***r* = .21 - .41** with SR social competence	***r* = .34 - .37** with personality scales ***r* = .08 - .14** with AI tasks
Level of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990; Lane, Sechrest, Riedel, Weldon, Kaszniak, & Schwartz, 1996; Ciarochi, Scott, Deane, & Heaven, 2003)	Emotional awareness		V	HO	Open response ratings	O / 20	AR	interrater (self/other): .84 / .92	***r* = -.11 - .07** with SR AI and AI , and TAS[4] ***r* = .43** with an emotion perception task[10]	-
total score								α = .88		
self		U / P						α = .84		
others		U / K						α = .83		
Empathic accuracy test (Ickes et al., 1990; Ickes, 2001) content accuracy valence accuracy	Empathic accuracy	U	(V)	HE	TS	O / M / -	G	interrater: .94 (content accuracy)	-	***r* = -.14 - .06** with personality traits[9]

Table 5.3 *continued*

Tacit Knowledge Inventory for Managers (TKIM; Wagner & Sternberg, 1991; Weis & Süß, 2007)	Tacit knowledge	K / U	V	HE	ECS	R / 91 (9*10 +1)	G	α=.74-.80 r_{retest}=.78 (3 weeks)	***r* = -.06 - .58** with TK-Tests (Gottfredson, 2003) ***r* = .21** with IPT-15[20] ***r* = .09** with Faces (MSCEIT)[20] ***r* = .36** with ST (4-Factor-Test)[20] ***r* = .07** with SIT[20]	***r* = .00 - .25** with AI scales (Gottfredson, (2003)
Interpersonal Perception Task – 15 (IPT-15; Costanzo & Archer, 1993; see also Weis & Süß, 2007)	Social perception	U	F	HS	SBS	M / 15	G	*KR-20*=.38 r_{retest}=.73	***r* = -.09** w. facial emotion perception[5] ***r* = .21** with TKIM[20] ***r* = .13** with Faces (MSCEIT)[20] ***r* = .09** with ST (4-Factor-Test)[20] ***r* = .15** with SIT[20]	***r* = -.02** with figural AI test[20] ***r* = -.19** with verbal AI test[20]
Wisdom-related knowledge test (Staudinger, Smith, & Baltes, 1994)	Wisdom-related knowledge	K	V	HE	SBS	O / -	AR	-	-	-
Situational Judgment Test of Social Intelligence (Legree, 1995; N = 193) Dinner-related knowledge Knowledge of indicators of alcohol abuse	Social insight / Tacit knowledge	K / U	V	HE	ECS	R / 20 each	AR	α = .50 α = .75	***r* =.34 / .25** with the U.S. Army SJT (dinner and alcohol abuse scale, respectively)	***r* = -.20 - .65** with AI scales
Emotion Accuracy Research Scale (Mayer & Geher, 1996; Geher, Warner, & Brown, 2001)	Emotion identification	U	V	HE	TS; GCS	R / 96 (8*12)	G	α_T = .24 α_G = .53	between scoring methods: ***r* =.14**[13] **/ .02**[8]	T / G-scoring : ***r* = .13 - .16 / .24 - .13** w. trait empathy[13] ***r* = -.06/.26** with AI[13]
Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto, LeRoux, Wilson-Cohn, Raroque, Kooken, Ekman, et al., 2000)	Emotion recognition	P	P	NO	GCS / SBS	RT / 56	AC	only forced choice: α=.82-.92 r_{retest}=.78 (3-4 weeks)	***r* = .50** with MERT, PONS[2]	only forced choice: ***r* = .31 - .38** with openness[12] ***r* = .15 - .35** w. conscient.[12] ***r* = -.45** with neuroticism[12]

Table 5.3 *continued*

Diagnostic Analysis of Nonverbal Accuracy Scale (DANVA2; Nowicki & Duke, 1994, 2001) adult facial expressions (AF) adult paralanguage (AP)	Nonverbal receptivity / emotional sensitivity	P / U	A P	NO	ECS	M / 24 M / 24	AC	α = .71-.78 α = .71-.78 / r_{retest}=.88	AF: ***r* =.54 - .58** with DANVA1-AF[15] AP: ***r* =.39 - .43** with SR social competence	-
Vocal Emotion Recognition Test (Vocal-I; Scherer, Banse, & Wallbott, 2001)	Emotion recognition	P	A	NO	SBS	M / 30	AC		Roberts, Schulze, O'Brian, MacCann, Reid, & Maul (2006)[17]: ***r* =-.10 - .24** with MSCEIT subscales (Faces - Blends) ***r* =-.10** with JACBART	***r* = .18** with g_f[17] ***r* =.20 - .25 / .00** with verbal / figural AI[17]
Facially Expressed Emotion Labeling (FEEL; Kessler, Bayerl, Deighton, & Traue, 2002)	Emotion perception	P	P	NO	SBS	M / 42	AC	α = .76 $r_{split\ half}$ =.73	-	-
Verbal Social-Cognitive Flexibility (SCF-V; Lee et al., 2002; N = 239)	Divergent thinking in social situations	C	V	HE	Open ans-wer ratings	O / -	AR	α = .88	***r* =.41** with SCF-P ***r* =.21 / .15** with social verbal / pictorial knowledge test	***r* =.35 / .31** with the verbal / pictorial AI creativity test
Pictorial Social-Cognitive Flexibility (SCF-P; Lee et al., 2002; N = 239)	Divergent thinking in social situations	C	P	HE	Open ans-wer ratings	O / -	AC	α = .89	***r* =.41** with SCF-V ***r* =.23 / .23** with social verbal / pictorial knowledge test	***r* =.27 / .24** with the verbal / figural AI creativity test
MSCEIT (Mayer et al., 2002) overall area: experiential (Branch I and II) area: strategic (Branch III and IV)						141		α = .90 α = .89 α = .84	***r* =.17-.53** among single MSCEIT-tests[14] overall EI: ***r* =.17 - .52** with SR EI and Empathy[14] ***r* =-.11 - .48** among single MSCEIT-tests[17] ***r* =-.02 - .20** of tasks with JACBART[17] ***r* =-.10 - .24** of tasks with VOCAL-I[17]	overall EI ***r* = .05** with figural reasoning (Raven)[14] ***r* = .36-.38** with verbal AI (Army Alpha)[14] ***r* = -.19 - .31** with personality scales[14]
Branch I: Perception: Faces	I: Emotion perception	P	P	NO	GCS / ECS	R / 20	AC	α = .82		
Pictures		-	P	NO		R / 30	AR	α = .85		
Branch II: Facilitation Facilitation	II: Using emotions	K	V	HE		R / 15	AR	α = .67		
Sensations		-	V	HE		R / 15	AR	α = .62		
Branch III: Understanding Changes	III: Emotional understanding	U	V	NO		M / 20	AR	α = .65		
Blends		U	V	NO		M / 12	AR	α = .52		
Branch IV: Regulation Em. Management	IV: Emotion regulation	K	V	HE		R / 20	AR	α = .78		
Em. i. Relationships		K	V	HE		R / 9	AR	α = .64		

Table 5.3 *continued*

Test of Emotional Intelligence (TEMINT; Schmidt-Atzert & Bühner, 2002; Amelang & Steinmayr, 2006)	Perception and understanding of emotions	U	V	HE	TS	R / 12	G	Schmidt-Atzert & Bühner (2002) / Amelang & Steinmayr, 2006) α=.76 / .77	-	***r* = .20 / .24 / .11** with reasoning / figural / verbal AI[18] ***r* = .22** with openness[18] ***r* =.11 - -.03** (with other Big Five) [18]
Facial Emotion Inspection Time Task (Austin, 2004) Happy IT Sad IT	Speed of emotional information processing	P	P	NO	RT	RT / 112 each	AC / AR	-	***r* =.42** between scoring procedures[1] ***r* =.40/.33** with Ekman-60 (Happy/Sad IT) [1] ***r* =-.14 - .25** with SR EI[1]	***r* = .48** with symbol inspection time task[1] ***r* = -.09/.07** with verbal AI (Happy/SadIT) [1]
Test for the Assessment of Empathy (Kunzmann & Richter, 2004)	Empathic accuracy	U	F	HS	TS / ECS	R / 8	G	-	-	-
Multimodal Emotion Recognition Test (MERT; Bänziger, 2005)	Emotion recognition	P	A / F	NO	SBS	M / 120	AC		***r* = .50** with JACBART, PONS[2]	
Test of Emotional Abilities (Freudenthaler & Neubauer, 2005; N = 277) [7] self (intrapersonal emotional abilities) others (interpersonal emotional abilities)	Emotion regulation in the self and others (typical performance)	K / U	V	HE	ECS	M / 18 M / 18	AR	α = .72 α = .70	**$r_{self/others}$=.29** ***r*=.08 - .18** (self with SR EI)[7] ***r*=-.03 - .08** (others with SR EI)	***r* = -.02 - .11** with AI (both scales) ***r* = .11 - .51** (self with personality) ***r* = -.08 - .35** (others with personality)
Situational Test of Emotional Understanding (STEU; MacCann, 2006; N = 207) [11]	Emotional understanding	U	V	HS	SBS / ECS / GCS	M / R 42	AR	S-scoring: α = .71	***r* = .42** with STEM ***r* = .32** with Stories (MEIS[h])	***r* = .49** with verbal AI ***r* = -.06 - .16** with personality scales

Table 5.3 *continued*

Situational Test of Emotion Management (STEM; MacCann, 2006; only ratings-based formats; N = 207)[11]	Emotional management	K / U	V	HS	ECS / GCS	M / R 44	G	GCS α = .86	***r* = .42** with STEU ***r* = .44** with Stories (MEIS)	***r* = .26** with verbal AI ***r* = -.09 - .31** with personality scales

Note. - no information available

a U = Understanding, M = Memory, P = Perception, C = Creativity, K = Knowledge

b V = based on written language, A = based on spoken language, P = based on pictures, F = based on video film

c HO = Homogeneous (restricted) context information, HE = Heterogeneous context information (not systematically varied), HS = Heterogeneous context information (systematically varied), NO = no context information provided

d SBS = standards-based scoring (one correct answer), RT = reaction times, TS = target scoring, GCS = group consensus scoring, ECS = expert consensus scoring

e MC = multiple choice, O = open response format, R = ratings-based scale, RT = reaction time

f G = genuine situations / persons, AC = actors, AR = artificially constructed

g SR = self-report data, AI = academic intelligence, EI = emotional intelligence, TK = tacit knowledge

h MEIS: Multifactor Emotional Intelligence Scale (Mayer, Caruso, & Salovey, 1999)

[1]Austin (2004): N = 72-92, [2]Bänziger (2005): N = 70, [3]Barnes and Sternberg (1990): N = 40, [4]Ciarrochi et al. (2003) : N = 331 , [5]Davies et al. (1998): N = 131, [6]Dymond (1949): N = 80, [7]Freudenthaler and Neubauer (2005): N = 277, [8]Geher et al. (2001) : N = 124 , [9]Ickes et al. (1990): N = 76, [10]Lane et al. (1996): N = 380, [11]MacCann (2006): N = 207, [12]Matsumoto et al. (2000): N = 89, [13]Mayer and Geher (1996): N = 321, [14]Mayer et al. (2002): N = 1297 – 1673, [15]Novicki and Duke (2001): N = 166, [16]Riggio et al. (1991): N = 112 – 171, [17]Roberts et al. (2006): N = 138, [18]Schmidt-Atzert and Bühner (2002): N = 117 (TEMINT with AI), N = 94 (TEMINT with personality traits), [19]Stricker and Rock (1990): N = 108 - 122, [20]Weis and Süß (2007): N = 118 (correlations within SI domain), N = 101 (correlations between AI and SI), [21]Wong et al. (1995): Study 1 N = 134, Study 2 N = 227

Overview of Test Classes

The tests sampled in the database are very heterogeneous in terms of formal and content-related attributes and in terms of the intended measurement constructs. With respect to the formal characteristics, several classes of tests can be identified within the database.

a) Cognitive ability tests in a narrow sense

Prototypical *cognitive ability tests* are, for example, the George Washington Social Intelligence Test (Moss et al., 1955), the Four Factor Test of Social Intelligence (O'Sullivan & Guilford, 1976), most of the scales of the MSCEIT (Mayer et al., 2002) or singular test approaches such as the IPT – 15 (Costanzo & Archer, 1993), the Couples Test (Barnes & Sternberg, 1989), the PONS (Rosenthal et al., 1979) and many more. The typical problem related to this test approach represents the identification of the correct answer when no standards-based scoring can be applied (e.g., the IPT-15 applies standards-based scores).

b) Situational Judgment Tests

Several *Situational Judgments Tests* are also represented: the TKIM (Wagner & Sternberg, 1991), the Test of Social Intelligence (Legree, 1995), the MSCEIT subtests Emotional Management and Emotions in Relationships (Mayer et al., 2002), the Test of Emotional Abilities (Neubauer & Freudenthaler, 2005), and the Situational Tests of Emotional Understanding and Management (MacCann, 2006). A specific variation of the Situational Judgment Test paradigm was introduced by Feudenthaler and Neubauer (2005) who instructed participants to indicate their typical response or behavior instead of the maximum performance (see Appendix A). However, it is questionable whether this type of assessment still measures cognitive performance because of the partly large correlations with personality traits. A more general problem associated with the Situational Judgment Tests is the lack of clarity about what is measured by this type of test (i.e., knowledge or inference). This should partly depend on the test takers knowledge in the queried content domain of one test.

c) Tests based on information processing paradigm

A separate class of tests represent those that are based either on the speed of information processing or, at least, on very basic processing demands such as unspeeded basic emotion perception. These are, for example, the JACBART (Matsumoto et al., 2000), the Vocal-I (Scherer et al., 2001), the FEEL (Kessler et al., 2002), and the Facial Emotion Inspection Time Tasks (Austin, 2004).

d) Test based on postdiction paradigm

The last broad category represents the so-called postdiction paradigm (O'Sullivan, 1983) that typically requires the identification of a mental state or personality trait of a target person. Tests classified into this category are, for example, the Dymond Rating Test (Dymond, 1949), the Empathic Accuracy Test (Ickes et al., 1990), the EARS (Mayer & Geher, 1996) or the TEMINT (Schmidt-Atzert & Bühner, 2002). The most striking problem of this test approach represents the question of the truthfulness or validity of the target's answer as the criterion for judging performance.

The empathic accuracy test by Ickes and colleagues (1990, see also Ickes, 2001) represents a tests that cannot be classified to the aforementioned approaches and is based on the interaction paradigm. This approach is based on genuine persons being put into an artificial encounter with other people. Every participant is judge and target at the same time. Therefore, they diverge from all other tests by enclosing the ego-involvement of the judge which is supposed to result in different forms of task requirements.

In summary, the intended measurement constructs as postulated by the test authors cover a broad range of social and emotional ability domains. The classifications of the tests into the performance model of Weis & Süß (2005) are done with care and sometimes were not possible. Particularly, knowledge and understanding requirements cannot be disentangled for the class of Situational Judgment Tests. Some of them explicitly claim to measure knowledge (e.g., TKIM), and some apply other labels (e.g., emotional management). Looking at the task requirements, it is hardly possible to completely exclude reasoning requirements from this type of test. The extent to which they influence performance should depend on the test taker's acquired knowledge (i.e., expertise) in this domain and the amount of available background information (Wagner & Sternberg, 1985). Finally, this represents an empirical question not addressed any further at this point.

Two Examples of Broad Test Batteries

Before integrating results concerning the validity of social intelligence tests, two prominent broad test batteries will be presented in more detail: the Four Factor Test of Social Intelligence (O'Sullivan & Guilford, 1976) as the only published broad test battery in the domain of social intelligence, and the MSCEIT as the most discussed measurement instrument in the literature on new ability constructs. Table 5.4 presents the scale descriptions and example items of the Four Factor Test which consists of four subtests, three based on paintings and cartoons, and one based on written language. The correct solutions are indicated by a cross in Table 5.4. Before this test was published, O'Sullivan et al. (1965) developed a large number of tasks intended to operationalize all 30 cells of the behavioral domain of the SOI. The stimuli of these tasks were photographs, drawings, cartoons, and tape-recorded words and sounds. Tasks were constructed analogously to the tasks of the other content facets of the SOI. To compose the final test battery, tasks were selected based on the best reliabilities and the highest factor saturations. Guilford and colleagues found support for this test approach from their first empirical validity studies. The authors reported no substantial correlations with general intellectual abilities (O'Sullivan & Guilford, 1966).

However, later studies showed more equivocal evidence. Probst (1975, 1982) also found empirical support for an independent ability construct of social intelligence. But factor analysis did not yield a common social intelligence factor comprising tests of different measurement approaches (e.g., self-report) or cognitive operations (e.g., understanding and memory). Results from Riggio et al. (1991) could neither support convergent nor divergent construct validity (i.e., with self-reported social intelligence and academic intelligence, respectively). In an exploratory factor analysis, the subscales of the Four Factor Test loaded on one factor with the Wechsler Adult Intelligence Scale – Revised Edition (WAIS-R, Vocabulary Subscale; Wechsler, 1981), showing near to zero correlations with the Social Skills Inventory (SSI; Riggio, 1989).

Table 5.4

The Four Factor Test of Social Intelligence: Test Descriptions and Examples

Four Factor Test of Social Intelligence (O'Sullivan & Guilford, 1976)

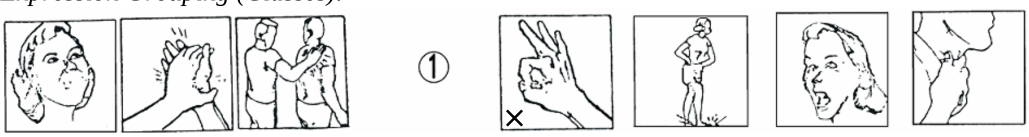
Test description

The test is intended to assess the behavioral cognition facet of the Structure of Intellect Model (Guilford, 1967) focusing on four factors of the product facet (i.e., respective product domain indicated in parentheses).

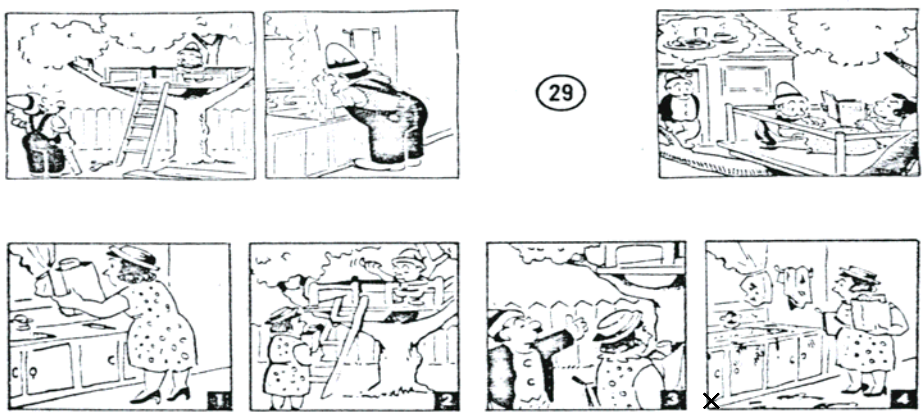
- *Expression Grouping (Classes):* Participants have to find one facial expression, out of four alternatives, which best fit a group of three other expressions.
- *Missing Cartoons (System):* Participants are required to fill-in a blank, in a sequence of cartoons, by selecting the correct cartoon out of four choice alternatives.
- *Social Translations (Transformations):* This test was the only written language measure of social intelligence in this battery. Participants are given a statement made between a pair of people, in a defined social relation. They have to choose one pair out of three alternatives that pair of people between whom the given statement has a different meaning.
- *Cartoon Prediction (Implications):* Participants are required to select one cartoon, out of three alternatives, that most appropriately adds to a given cartoon.

Examples

Expression Grouping (Classes):

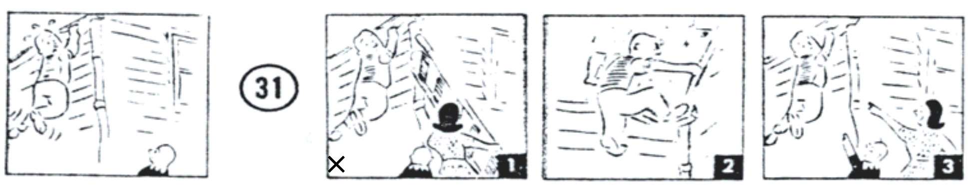


Missing Cartoons (System):



Social Translations (Transformations): Parent to a child: "I don't think so." Alternatives:
a. teacher to student, b. student to teacher, c. student to student.

Cartoon Prediction (Implications):



The analytical approach for the test construction was criticized by Probst (1973, 1982) who, conforming with Thorndike's original idea, questioned the content validity of these artificially produced tasks. Instead, Probst called for the operationalization of relevant social elements. Furthermore, he attributed the lack of validity to the concept underlying the scoring of the tasks. The scoring key was developed by a group of experts by inducing a consensus about the logical true and false solutions. Probst claimed that this expert consensus could only reflect a certain probability or plausibility value of different response alternatives. It cannot reflect "true or false answers in a logical sense" because these "stimuli are not interpretable without the situational context" (Probst, 1982, p. 221). Context and background information eventually provide meaning to certain social cues and a lack of this causes equivocity, which was not accomplished by the Four Factor Test. Apart from Probst's critique, other more specific methodological concerns refer to the imbalance of the applied task material (i.e., one is verbal, three are nonverbal). Moreover, the stimuli clearly lack face validity and more importantly, the present day opportunities of digital stimulus presentation to enhance real-life fidelity must not be neglected today.

The MSCEIT represents a more contemporary approach to develop a cognitive ability test for the domain of social and emotional abilities. The MSCEIT is based on the Four-Branch-Model of Emotional Intelligence (Mayer & Salovey, 1997). Table 5.5 presents the scale descriptions and examples for most of the subscales. The manual reported results from exploratory factor analyses that supported the postulated one-, two-, and four-factor solutions of the Four-Branch-Model (i.e., general emotional intelligence, experiential vs. strategic emotional intelligence, and the Four Branches). Later, confirmatory factor analysis supported the four-factor solution with, assumably, uncorrelated factors (Mayer et al., 2003). The authors did not explicitly speak of uncorrelated factors but also did not report any factor intercorrelations. Roberts et al. (2006) conducted a confirmatory factor analysis with the MSCEIT, the JACBART (Matsumoto et al., 2000), the Vocal-I (Scherer et al., 2001), and measures of fluid and crystallized intelligence. They identified a three-factor solution with Factor 1 equaling experiential emotional intelligence (i.e., Branch I and II), Factor 2 strategic emotional intelligence (i.e., Branch II and IV), and Factor 3 showing loadings of the intelligence and the emotion measures of the JACBART and Vocal-I. The factor intercorrelations ranged from .28 to .34.

Although never particularly highlighted by the test authors, across studies the MSCEIT subscales showed substantial correlations with verbal academic intelligence that sometimes exceeded those between the MSCEIT subscales (see Table 5.3).

Table 5.5

The MSCEIT (Mayer et al., 2002): Test Descriptions and Examples

Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)
(Mayer et al., 2002)
Test description
The MSCEIT was developed to operationalize the Four Branch Model of Mayer & Salovey (1997). For each Branch, two tasks were constructed. The MSCEIT allows the score building for a general emotional intelligence factor, the Four Branches, and additionally, for two area scores: experiential emotional intelligence (Branch I and II) and strategic emotional intelligence (Branch III and IV).
Branch I: Emotion Perception
Faces (Section A): Test takers are presented the pictures of four faces of persons that display ambiguous emotion expressions. They have to rate how much a feeling is expressed in the face on a 5-point scale (from no feeling to extreme feeling expressed).
Pictures (Section E): Test takers are presented six pictures of landscapes and pieces of art. They have to rate how much a feeling is expressed in the pictures a 5-point scale.
Branch II: Emotion Facilitation of Thought
Facilitation (Section B): Test takers are presented 5 short written situation descriptions and have rate the usefulness of three different emotions for dealing with that situation on a 5-point scale (from not useful to useful).
Sensation (Section F): Test takers read five descriptions of the emotional / sensational life of a fictitious person. They have to indicate for each of three alternative sensations / emotions how much the aforementioned emotion or sensation is alike to the respective sensations or emotions on a 5-point scale (from not alike to very much alike).
Branch III: Emotion Understanding
Changes (Section C): Test takers read 20 brief written descriptions of the emotional life of a person in a specific situation. They have to complete the last sentence by selecting one of five alternative emotions which best complements the description.
Blends (Section G): For twelve items, test takers have to select out of five alternatives the best combination of emotions for a given feeling.
Branch IV: Emotion Regulation
Emotion Management (Section D): Test takers read five descriptions of situations ended up by a goal. For each situation, four alternative actions have to be rated according to the effectiveness to accomplish the goal on a 5-point scale (from very ineffective to very effective).
Emotions in Relationships (Section H): Test takers are presented three situation descriptions that include special interpersonal problems. Three alternative actions have to be rated on a 5-point scale in terms of their effectiveness in solving the problem.

Table 5.5 *continued*

Examples (fictitious)
Branch II: Emotion Facilitation of Thought
Facilitation (Section B): What mood might be helpful to feel when preparing for an exam? Emotions rated: joy, tension, frustration
Sensations (Section F): Imagine you feel guilty for having forgotten your mother's birthday. How much is the feeling like each of the following? (a. salty, b. blue, c. cold)
Branch III: Emotion Understanding
Changes (Section C): Johnny felt full of frustration and discontent. He then thought about what he could do in the future and began to feel ... (a. excited, b. satisfied, c. angry, d. depressed, e. surprised).
Blends (Section G): A feeling of being bothered often combines the emotions of ... (a. love, surprise, embarrassment, b. anger, fear, amazement, ...)
Branch IV: Emotion Regulation
Emotion Management (Section D): Maria went to bed and thought about her day. She could not stop thinking about some bad things that have happened. How effective is each of the following actions in helping her fall asleep?
Action 1: Maria started to think about what beautiful things she could do the next day.
Emotions in Relationships (Section H): Carol was very engaged in work-related problems and couldn't afford losing time. Then, her best friend asked her to help her with renovating her apartment. Last time, her friend had asked for help, she already had to resign because of her work. How effective are the following actions in maintaining a good relationship to Carol?
Action 1: Carol tried to find a way to delay some important tasks at work so that she could spare time to help her friend.

With regard to the incremental predictive validity for relevant external criteria, the results are equivocal. Burns, Bastian, and Nettelbeck (2007) criticized criterion validity studies for not controlling for either academic intelligence or personality as established predictors or both. When demographic variables, personality, and intelligence were controlled for, the incremental validity was weak (Burns et al., 2007; Brackett & Mayer, 2003). Brody (2004) also reviewed criterion validity studies and came to the same conclusions. Moreover, he criticized that sometimes, the general emotional intelligence factor was applied for prediction, sometimes, scores representing one of the Four Branches. Brody required a Schmid-Leiman solution of the factor-structure of the MSCEIT that allows the modeling of general and more specific factors at one time.

Besides the lack of empirical support for the factor structure and high correlations with verbal academic intelligence, the pivotal criticisms address the scoring and the cognitive requirements of the single tasks and their match with the Four-Branch-Model. Above all, the scoring by group consensus suggests no objective testing and was already discussed in Chapter 5.2.4. This put aside, tasks for Branch I (i.e., *ability to perceive emotions*) are restricted to emotion expressions in faces, landscapes, and artificial pictures. Further components of Branch I such as the perception of emotions in oneself and appraisal of emotions in oneself and others are not included in the MSCEIT. Tasks of Branch III (Understanding Emotions) are also restricted to emotions in others. Branch II and IV are represented by more complex tasks including situation-related information in the stimuli. The necessary requirements to accomplish these tasks are basically knowledge (Mayer et al., 2003; see also Brody, 2004) while the Branch definitions demand behavioral requirements. The effect of method-related variance on the covariance structure within the MSCEIT and between the MSCEIT tasks and other measures of academic intelligence and emotions is not discussed by the test authors. MacCann (2006) shows that correlations are lower for tasks that use different response formats. The use of rating-based scales in the MSCEIT (except for *Blends, Section G*) may still affect the size of correlation within the MSCEIT and with academic intelligence tests that are typically not based on rating-scales.

The two test batteries show substantial weaknesses. From today's perspective, if they cannot serve as instruments to assess social or emotional intelligence, they may not as well serve as instruments to validate any new measure in this research domain.

5.4 Integrating Findings about the Validity of Social Intelligence

5.4.1 Empirical findings

Referring to the test database in Table 5.3, a predominant number of tests turn out to be based on written task material (i.e., 23 out of 47 tests or subtests). Another 12 tests are based on pictures. Obviously, the most genuine and relevant task material (i.e., auditory and video-based) are totally underrepresented. Remarkably, just four tests systematically vary situative or contextual information. However, the underlying taxonomy is sometimes very simple and restricted to one or two varied dimensions with a small number of elements (e.g., positive vs. negative reported emotional experiences in the Test for the Assessment of Empathy, Kunzmann & Richter, 2004).

The reliability coefficients are generally satisfactory, however, for some tests, the reported range of reliability is large (e.g., the Four Factor Tests, MSCEIT subscales). For only five tests, the convergent construct validity coefficients exceeded those for divergent construct validity. These are, on the one hand, two Situational Judgment Tests (the TKIM and the Situational Test of Emotional Understanding), and, on the other hand, three tests based on pictorial or video material (the CARAT, the Couples Test, and the Facial Emotion Inspection Time Task). In general, evidence for the convergent construct validity is weak (Ambady, LaPlante, & Johnson, 2001; Buck, 1983; Hall, 2001). Ambady and colleagues attributed this to the differences in the queried modality (e.g., emotions, cognitions, personality, etc.) and the diverse types of stimuli (i.e., posed vs. genuine vs. artificial; presented in language, voice, face or body). Comparably, variations in the divergent construct validity can be attributed to the differences in the applied measurement approaches in terms of the formal characteristics (e.g., item formats and scoring), the applied task material (i.e., above all written language-based tests), and the applied validation instruments (i.e., verbal vs. figural-spatial academic intelligence tests).

However, all the present results are based on univariate analyses. Only a few studies applied multivariate designs or, at least, multiple measures of social intelligence, to assess the construct validity (Barchard, 2003; Buck, 1983; Davies et al., 1998; Hall, 2001; Keating, 1978; Lee et al. 2000; Lee et al., 2002; Roberts et al., 2006; Weis & Süß, 2007; Wong et al., 1995). Keating (1978) applied three written language indicators of social intelligence (e.g., Chapin Social Insight Test, Chapin 1967, Gough, 1968) besides three measures of academic intelligence (both verbal and nonverbal material). Neither correlational nor factor analytic results supported construct validity. Within-domain correlations did not exceed across-domain correlations, and no coherent factor structure could be identified. In a comprehensive study, Barchard (2003) applied the MSCEIT and the Four Factor Test of Social Intelligence both as indicators for emotional intelligence. Surprisingly, she did not report within-domain correlations or factor analytic results about convergent validity evidence but aggregated the tests to one scale of emotional intelligence.

During the last decade, multitrait-multimethod (MTMM; Campbell & Fiske, 1959) designs and analysis techniques were applied to investigating the construct validity of social intelligence (Lee et al. 2000; Lee et al., 2002; Weis & Süß, 2007; Wong et al., 1995). Except for Weis and Süß (2007), all these investigations applied written language and pictorial performance tests besides self-report inventories for the assessment of the respective ability factors. Weis & Süß included video-based performance tests. The use of confirmatory factor analysis in all studies allowed the separation of ability- and method-related variance. Wong et al. (1995, Study 1) assessed academic intelligence, social perception, and socially intelligent behavior. Written language social perception was operationalized by the test *Recognition of the Mental State Behind Words* (Moss et al., 1955). The test *Expression Grouping* (O'Sullivan & Guilford, 1976) should measure nonverbal social perception. Results yielded a model with four uncorrelated method-factors (i.e., written language, nonverbal, self-report, and other-report) and three correlated ability-factors (i.e., academic intelligence, social perception, and heterosexual interaction). However, the correlative and factor-analytic results suggested a substantial overlap between social perception and academic intelligence (r = .67 on a latent level), which exceeded the correlation of social perception with socially effective behavior (r = .54). Wong et al. (Study 2) assessed social perception, social insight, and social knowledge. The test *Social Translation* served as a measure for written language social perception (O'Sullivan & Guilford, 1976), *Expression Grouping* as the nonverbal test. Social insight (written language) was operationalized by the *Judgment in Social Situations* test (Moss et al., 1955), social insight (nonverbal) by the test *Cartoon Prediction* (O'Sullivan and Guilford, 1976). Confirmatory factor analysis successfully identified the cognitive ability-factors of social insight and social knowledge separable from but positively correlated with academic intelligence. However, social perception could not be separated from social insight. Lee et al. (2002) operationalized social knowledge and the social-cognitive flexibility (see Table 5.3 and Appendix A for a description of the measures of social-cognitive flexibility). Results of this study showed separable social intelligence ability factors distinct from but positively correlated with (general) creativity.

In summary, the just described MTMM studies brought clear evidence for the multidimensionality of social intelligence and, in part, for the divergent and convergent construct validity. But still, some criticisms remain. *(a)* The assignment of the single tests to the ability domains in the different studies did not follow a consistent principle. For example, tests assigned to measure social perception were later applied as tests of social insight. *(b)* The factor loadings on the ability factors varied between performance measures and self-report inventories, although the method-related variance of self-report data was controlled for by the introduction of method-related factors. Therewith, the influence of self-report data on the identified factor structure could not be clarified. Thus, a pure performance construct could not be supported by data. *(c)* Moreover, no further convergent validity evidence was available since self-report data were already included in the social intelligence models.

Weis and Süß (2007) applied a MTMM design that operationalized social understanding, social memory, and social knowledge. Written language, pictorial, and video-based tests were applied. The written language measures of social understanding were the *Chapin Social Insight Test* and the *Social Translation Test*. The pictorial measure of social understanding was the *Faces Test* (Branch I in the MSCEIT; Mayer et al., 2002), and the video-based measure the *IPT–15*. Tasks for social memory were all newly constructed. The

TKIM served as the written language measure of social knowledge. Confirmatory factor analysis supported the postulated structure of correlated ability factors (i.e., social understanding, memory, and knowledge) and of a general social intelligence factor. In both models, method variance related to indicators based on written language was controlled by the introduction of a written language method factor. In the three-factor solution, social knowledge correlated significantly with social memory and social understanding (.42 and .50, respectively). Social memory and social understanding also correlated significantly (i.e., .45). The loadings on the written language method factor were heterogeneous, but all indicators loaded positively on this factor. Correlational and multiple regression analysis showed domain specific overlap of the social intelligence ability factors with specific domains of the BIS-Test (Jäger et al., 1997).

In extension to previous MTMM studies, the factor structure was independent from self-report data so that a multidimensional performance construct of social intelligence was supported. The study of Weis and Süß (2007) also brought clear evidence for a bias of method-related variance of tests based on written language. Moreover, the study demonstrated the importance of MTMM designs and particularly, the application of auditory and nonverbal measures.

5.4.2 Critical Summary

To conclude the present chapter, the most striking methodological problems surrounding the measurement of social intelligence will be summarized and integrated.

a) Lack of reference to theoretical models and conceptual confusion

Tests or test batteries based on a theoretical model are sparse. Tests covering the same measurement constructs were given diverse labels (e.g., tests of social understanding can be found under the labels of nonverbal decoding skills, interpersonal sensitivity, empathic accuracy, and receiving abilities). Tests with identical labels sometimes assess different constructs (e.g., tests of social perception sometimes include interpretative requirements). Hence, attempts to integrate research in the field of social and emotional abilities first need to disentangle the different approaches so that the research can be interpreted profoundly. In this context, Bernieri (2001) demanded a univocal localization of the measurement construct within related construct “to avoid overgeneralization and misinterpretation” (p. 8).

b) Unsystematic application of measurement approaches

Social intelligence literature includes diverse measurement approaches, all applied to measure the same construct. On the most general level, measurement approaches can be subdivided into assessing T-data (Cattell, 1965; i.e., cognitive ability tests), Q-data (Cattell, 1965; i.e., self-report questionnaires), and behavior. Approaches can additionally be classified in terms of the applied material (or the task contents, i.e., written and spoken language, pictures, videos), the complexity and broadness of stimuli (e.g., the amount of context information), the queried modality (e.g., emotions, cognitions, etc.), item formats, and scoring procedures. The different approaches, however, were applied unsystematically within and across empirical studies. Thus, no consistent pattern of validity results emerged from empirical investigations. Again, the interpretation of results suffered from a lack of systematic and methodologically elaborate research designs. Recent studies based on MTMM designs, however, could finally improve the informative value about construct validity.

c) Lack of genuineness and social relevance in task material

Only about half of the tests and subtests made use of auditory, picture- and video-based stimuli (24 out of 47). The relevance of the different communication channels in terms of their proportional use, however, seems to paint a different picture. According to Mehrabian and Ferris (1967), 93 % of communicative cues are transferred via auditory and visual channels. Only 7 % of relevant cues are transmitted over language contents. In addition, artificially constructed or posed task materials are no adequate substitution for the diversity of social situations that occur in real life. Thus, it seems inevitable to account for the genuineness and relevance of task material in order to validly assess social intelligence. This involves a balance of task material that addresses different communication channels and the systematic variation of task contexts that reflect real life requirements.

d) Decontextualized stimuli

Only a small number of tests included systematically varied context information and a large number did not include any. The application of decontextualized stimuli is not only problematic in terms of a lack of genuineness. Additionally, many of the described tests include requirements which cannot at all or hardly be accomplished without context information or situative cues. This is especially true for complex stimuli that allow several alternative interpretations. If tests provided complementary context information, they were rarely based on taxonomic considerations so that the generalizability of test results is limited.

5.5 Social and Emotional Intelligence – Their Qualification and Value as a New Intelligence Construct

To wrap up and conclude the theoretical and methodological foundations, social and emotional intelligence shall be evaluated against the formerly established requirements for new intelligence constructs. The basis for this evaluation is the knowledge accumulated throughout this thesis and up to the present chapter. This refers to the knowledge of the modified performance model of social intelligence, the Four-Branch-Model of Emotional Intelligence, and current measurement approaches. Table 5.6 summarizes the results of this evaluation in a simplified manner.

Table 5.6

Social and Emotional Intelligence Compared to the Requirements for Intelligence Constructs

	Requirement	**Social Intelligence**	**Emotional Intelligence**
Theory development	A-priori theoretical considerations about the localization of the intended construct within the sphere of individual differences (Matthews et al., 2005)	+	-
	Clear and nonredundant terminology (O'Sullivan, 1983)	+	-
	Based on empirical results (Süß, 2001)	+	-
Construct specification	Definitions of the underlying cognitive requirements (Carroll, 1993)	+	-
	High generality (heterogeneous in terms of contents) (Süß, 2001)	+	+
	Stable over time (Süß, 2001)	∅	∅
	Minimum amount of knowledge requirements (Süß, 2001)	+	-
Operationalization	Objective T-data (Süß, 2001; Weber & Westmeyer, 2001)	(+)	(+)
	Objective scoring rules (Weber & Westmeyer, 2001)	o	(+)
	Psychometrically sound (Matthews et al., 2005; Weber & Westmeyer, 2001)	o	o
Validation	Convergent and divergent construct validity (O'Sullivan, 1983; Weber & Westmeyer, 2001)	o	-
	Careful selection of validation instruments (Schaie, 2001)	(+)	(+)
	Incremental predictive validity for heterogeneous external criteria (Süß, 2001)	-	-

Note. + accomplished
(+) accomplished only for some measurement instruments or studies
\- not accomplished yet
o equivocal results
∅ no information available

The most striking weakness of emotional intelligence is the lack of a coherent theoretical model with a clear terminology about the underlying ability domains and a statement about the role of emotions, cognitions, and knowledge. Moreover, no attempts are made to overcome this weakness and to advance the development of more elaborate theoretical models. Scherer (2007), however, did make a first attempt in sorting the concepts, but he refrained from calling it an intelligence. In the field of social intelligence, integrative approaches to form theories and clarify concepts were only recently commenced (Süß et al., 2005; Weis & Süß, 2005; Weis et al., 2006). Empirical findings could already provide support for the performance model of social intelligence. Empirical support for the Four-Branch-Model cannot be derived from the literature since construct definitions and measurement constructs do not match.

Operationalizations on both sides are not sufficiently mature. In terms of social intelligence, test approaches are too diverse to allow a clear statement about the objectivity and psychometric quality of the construct. For emotional intelligence, the MSCEIT alone does not allow objective measurement. Recent test developments (MacCann, 2006) and tests in the context of emotion research, however, show promising psychometric quality and, sometimes, allow objective scoring. In general, the psychometric properties of test approaches on both sides, in turn, vary substantially across studies and samples. Evidence for the divergent validity of tests of social intelligence is equivocal which was attributed to an unsystematic application of measurement approaches and a lack of MTMM designs. Evidence for the convergent construct validity was generally weak. Existing tests of emotional intelligence could not prove their divergent and convergent construct validity (i.e., high correlations with verbal academic intelligence and low correlations within the MSCEIT or with tests of emotion research). Moreover, a systematic variation task material in MTMM designs is not possible with the existing tests, which mostly consist of verbal task material. In summary, social intelligence can rely on an elaborated theoretical foundation, seemingly ahead of those of emotional intelligence. It was attempted to overcome the most central methodological problems in the subsequently described test development.

6 Program of the Present Work

Aims, Design, Test Construction Principles, and Research Questions

6.1 Aims and Objectives

The aims of the present work were twofold. The first aim dealt with test development including the processes and procedures that are involved in test development and the principles of test construction. The result of the test development would be a comprehensive test battery, the Social Intelligence Test Magdeburg (SIM), as called hereafter when the entire battery is addressed. The second aim addressed the construct validation of social intelligence as assessed by the SIM.

(1) Test development

a) Test development was based on the modified performance model of Weis and Süß (2005) which defines social intelligence as a heterogeneous and multidimensional performance construct. Single tasks should be designed such that they match the performance determinants of ability domains and contain only cognitive requirements with a minimum amount of knowledge.

b) By systematically varying the task material (i.e., the contents), a MTMM design was postulated which was directed at controlling for method-related variance in the single tasks.

c) Test development also took into account additional taxonomic considerations establishing a hypothetical faceted model of social intelligence. The taxonomy should systematically vary (a) the task contents related to different task material (i.e., the content of the tasks), (b) the context information, and (c) the queried modalities. Whether the postulated facets represented separable and meaningful ability domains was to be explored.

d) Task material solely comprised of genuine situations and persons. Therewith, situative cues and context information could be introduced.

e) Test development aimed at resulting in psychometrically sound scales in terms of psychometric item and scale properties. The effect of item and response formats on the psychometric properties was to be explored.

(2) Construct validation

The present work postulated to provide evidence for the construct validity of social intelligence as assessed by the SIM. On the one hand, the internal structure of social intelligence was to be examined. On the other hand, the relationship of social intelligence to academic intelligence and to personality traits was to be investigated in order to prove divergent construct validity.

6.2 Overall Design and Test Development

6.2.1 General Test Construction Principles

The MTMM design for test development contained three operative ability domains (i.e., *social understanding*, SU; *social memory*, SM; and *social perception*, SP). Social creativity and knowledge were omitted from the design. The definition of *social creativity* is not as explicit as adequate testing would require. In addition, the development of a scoring key to judge the number and diversity of “adequate” answers to a social problem would require a rule for the definition of what is “adequate” and a rule to determine diversity (i.e., what are meaningful categories). This is beyond the scope of the present work. The construction of an adequate *social knowledge* test seems an even more insurmountable problem. Social knowledge is context- and situation-specific. Test developers have to select the requested knowledge domain and, a-priori, determine what is correct or incorrect. The present aim was to assess a generally valid intelligence construct which stands in contradiction with the context-specificity of knowledge. Nevertheless, both social creativity and knowledge are interesting and worth pursuing in order to complement the model of social intelligence. Possible future directions in this concern will be addressed in Chapter 10.

The remaining three ability domains were operationalized by the use of written and spoken language (V = written language and A = auditory / spoken language, respectively), pictures (P), and videos (F = film). First and foremost, these four material-related content domains were intended to control for method-related variance within the single tasks. Except for one single occasion (see the test description of social understanding), audio and pictorial information were strictly separated. The final design of the SIM resulted in a 3 x 4 cross-classification of operations and contents illustrated in Table 6.1. This design represents the most basic classification common to all developed tasks. Initially, it was the intention to develop two tasks per cell. This goal was not accomplished within a single stage during the test construction process. The present work comprised several small pilot studies and two larger main studies. The complete SIM was applied only in the second main study (see Table 6.6 for an overview of the tasks applied in the studies).

Table 6.1

MTMM Design of the SIM Cross-Classifying Operations and Contents

	Contents			
Operations	**written language (V)**	**spoken language (A)**	**pictures (P)**	**videos (F)**
social understanding (SU)	SUv	SUa	SUp	SUf
social memory (SM)	SMv	SMa	SMp	SMf
social perception (SP)	SPv	SPa	SPp	SPf

Apart from the core MTMM design, the development of every task was based on a *simple taxonomy* that only accounted for *some* of the previously mentioned taxonomic principles of social situations (see Chapter 4.3.3). This taxonomy guided the recording and final selection of task material which was sampled with genuine persons within their real-life contexts. The stimulus material varied the number of persons displayed (i.e., just one person, a dyad, or small groups) and the setting (i.e., private settings vs. public settings). Private settings included those involving family, friends, and acquaintances. Public settings included occupational contexts (e.g., at work place) and public life in general (e.g., practicing in a sports club, shopping in public malls or visiting a public exhibition, etc.). As far as possible, the setting was also varied in terms of the prevalent topic on the two core dimensions of the *Interpersonal Circumplex* (i.e., love and power; Wiggins, 1979). Situations could differ depending upon whether people were engaged in interactions that dealt with *(a)* closeness vs. distance or *(b)* dominance vs. submission. The taxonomic variations could, in some cases, not be executed stringently *within* one task. It was, however, assumed that the intended effect was accomplished when the context was varied *across* tasks.

6.2.2 Test Construction Principles Underlying the Single Ability Domains

The following passages will address the basic test approach for each operative ability domain and the core elements of the test construction process.

6.2.2.1 Social Understanding (SU)

The basic test idea underlying the tasks of *social understanding* integrated the postdiction paradigm with a scenario approach. The scenario approach equaled the Situational Judgment Test paradigm in terms of compiling several items to one stimulus. Basically, two trains of thought were responsible for this decision. The first referred to the work with targets which is characteristic to the postdiction paradigm. It requires the test takers to judge the target in terms of different modalities (e.g., emotions, cognitions, relationships, personality traits, etc.). Performance is typically judged in terms of the deviation from the target's answer. The second train of thought referred to the availability of relevant contextual information for every item, but , at the same time, considered 'economic' testing. Providing contextual information for every single item for several targets that are not associated with each other would require a large amount of testing time. A scenario approach allows the embedding of several items in the situative context of one scene that provides *(a)* background information about the target and *(b)* situative cues.

The general procedure of all scenarios included the following elements:

a) One scenario was related to one target person (i.e., a stranger).

b) Each scenario started with an introduction of the target person himself or herself in a short video clip involving audio-visual information. It typically involved information about the biography of the targets including name, age, profession, hobbies, interests, and anything the target felt comfortable talking about. The introduction aims to provide test takers with an impression of the target's voice and physical appearance. Additionally, subjects were provided with written information about the target's biography which enframed the video-based self-presentation. A typical written introduction at the beginning of a scenario was as follows:

> *"Christoph is standing in the center of the upcoming tasks. Christoph is 23 years old and single. He studies [… some biographical information …]. You will now watch a video clip of Christoph introducing himself. Pay careful attention to how he looks and to the way he speaks. [… video clip …]. You now have an impression of Christoph. Now, turn to the first scene and the first task."*

c) Each scenario consisted of a minimum number of eight stimuli entities or scenes that represented the smallest structural elements. Each scene was introduced by a short text that described the upcoming scene and announced upcoming questions before the scene was played. This primed the test taker so that his or her attention could be directed towards relevant cues within the upcoming scene.

"You will now watch a video clip of Christoph and his girlfriend going swimming at a lake together with a group of friends. You will have to judge how Chistoph feels while saying hello to the group, what he wants to express by pulling his girlfriend into the water, and how often he is in a comparable situation with his girlfriend."

In Study 1, the test takers were given the option of reading over the items before being presented with the stimulus. This resulted in a much longer testing time and was therefore substituted with the introductory texts.

d) In order to accomplish the taxonomic demands, each type of setting (i.e., private vs. public) was represented by each type of task material (i.e., written and spoken language, pictures, and videos). The resultant were the above-mentioned eight stimuli scenes per scenario.

Target Selection and Material Sampling Process

The present work attempted to select heterogeneous targets in order to include as many diverse persons and situations as possible. This would help ensure a balanced impact of indiosyncracy (Buck, 1983; Sabatelli et al., 1980, 1982) and the judge-target similarity (Cline, 1964; Cronbach, 1955) on judgments, although, the placement of additional statistical controls would be valuable. Targets would equally represent both genders, different age group, different educational backgrounds (e.g., from high school graduates to university graduates), and different professions. Targets were selected from among family, friends, and acquaintances of the work group and came from different regions across Germany (e.g., from Berlin to Aachen and Bremen to Nürnberg). All potential targets to whom requests were sent, agreed to participate in the study. They were briefed about the aims of the study, the prospective efforts of material sampling, and the detailed process (see Appendix B for the original letter to the targets). Additionally, each other person involved in the recordings was also informed about the aims and the process of the study. After all participants had given their consent (with his or her signature) to being recorded and to releasing the recordings for scientific purposes, the targets were accompanied by a member of our staff (called the investigator hereafter) over a time period of two or three days. The recordings took place during the targets' typical everyday life. Additionally, the targets were asked to provide genuine stimulus material based on written language (e.g., email or mail correspondence, diary entries, etc). All investigators reported that the recordings did not seem to have an impact on people's behavior over the long run (i.e., after the first few minutes of recording). This corresponded with the findings of Carpenter and Merkel (1988) who could show that different types of observation (i.e., one-way mirror, audio recordings, and video recordings) did not result in different interactions between couples.

After a recording session, the material was viewed by both, the investigator and the target. While doing so, the target answered questions about his or her mental state with regard to specific scenes (assigned to an exact point in time). The targets responded either in open-ended formats or on a visual analogous rating scale where the endpoints were provided by the investigator. The analogous scale was 10 cm long. Sometimes, additional background information about the persons involved (e.g., type of relationship with the target, for example, siblings, friends for ten years, etc.) or the context of the situation (e.g., negotiations about payments that have been going on for a couple of weeks without any outcome, etc.) were sampled. Targets were also asked to fill out several questionnaires: *(a)* the NEO FFI (Borkenau & Ostendorf, 1993) as an inventory of the Big Five personality traits, *(b)* the Inventory of Interpersonal Problems-Circumplex (IIP-C; Horowitz, Strauss, & Kordy, 2000) as a measure of interpersonally relevant personality traits, *(c)* a social behavioral questionnaire based on prototypical acts (Amelang et al., 1989), *(d)* the social desirability scale of the Freiburger Persönlichkeitsinventar – Revidierte Fassung (FPI-R; Fahrenberg, Hampel, & Selg, 2001), and *(e)* a biographical questionnaire.

In the first large study, only four scenarios were implemented. Four additional scenarios were added after the first study so that the final SIM version contained eight scenarios related to eight target persons. Table 6.2 shows the demographic characteristics of the final eight targets in alphabetical order and indicates the study assignment. Figure 6.1 and 6.2 show the targets' personality profiles on the Big Five and on the IIP-C. The scales range from 0 to 4, a higher score is indicative of a higher degree of the respective trait. The first four targets (Study 1) were restricted to a certain score and age range. Obviously, the heterogeneity of the profiles in terms of the demographic and personality variables was enhanced from Study 1 to Study 2.

Table 6.2

Demographic Characteristics of Target Persons in Alphabetical Order

target name	**gender**	**age**	**profession**	**Education U / H**	**Study 1/ 2**
Bringfried	male	43	medical doctor	U	2
Christoph	male	23	student of law	U	1 / 2
Conny	female	41	owns a bistro	H	2
Friedrich	male	69	estate agent	H	2
Hannah	female	60	teacher / social education worker	U	2
Katharina	female	26	student of psychology	U	1 / 2
Matthias	male	33	dancing teacher (owns school)	H	1 / 2
Renate	female	24	biotechnical assistant	H	1 / 2

Note. U = university education, H = highschool education
Study 1 / 2 = applied in Study 1 and 2, Study 2 = applied only in Study 2

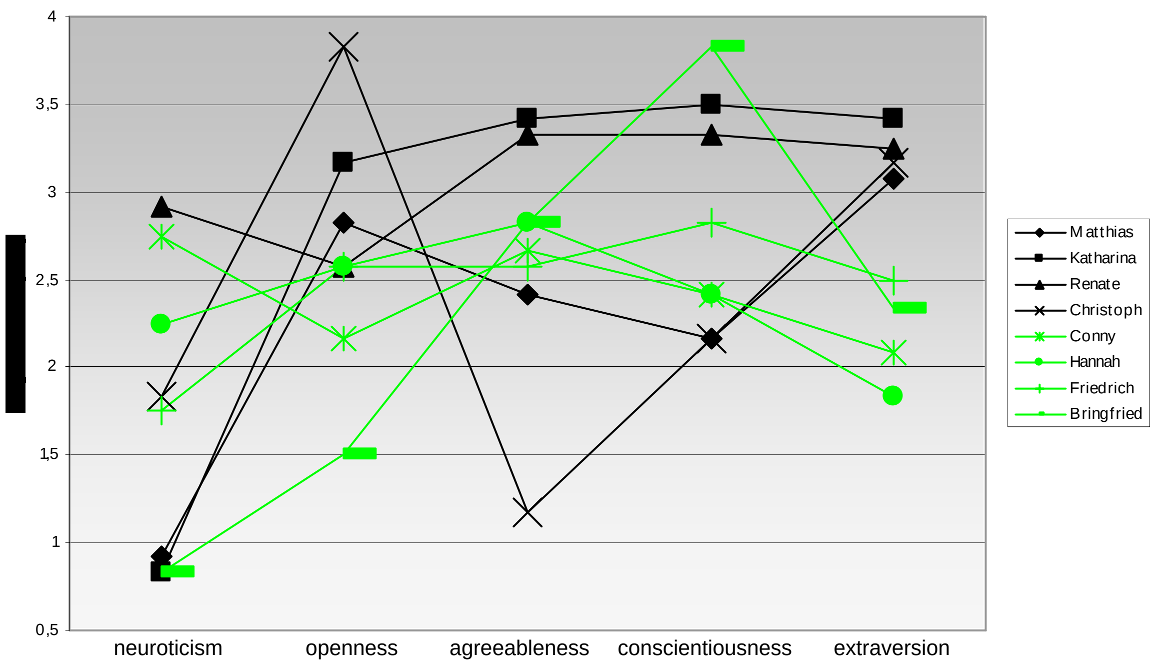


Figure 6.1

Personality Profiles of the Targets on the Big Five (NEO-FFI, Borkenau & Ostendorf, 1993)

Note. black lines = Study 1 / 2, green lines = only Study 2, a high score on a particular trait indicates high expression of that trait

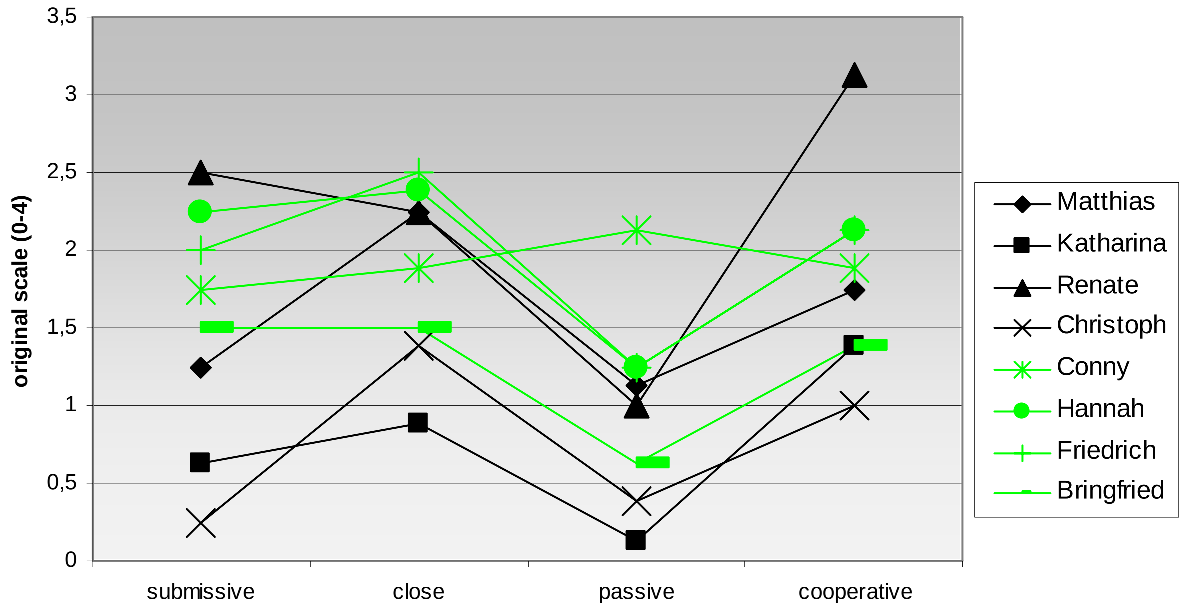


Figure 6.2

Personality Profiles of the Targets on the Interpersonal Circumplex (IIP-C, Horowitz et al., 2000)

Note. black lines = Study 1 / 2, green lines = only Study 2, a high score on a particular trait indicates high expression of that trait

Item Selection

The recorded material was selected and edited so that adequate stimuli emerged. This basically implied selecting those entities within the recorded material that presented interesting and relevant social situations in response to which target information was available. After identifying these entities, the material was edited (shortened) so that relevant elements were still included but stimulus length reduced. Care was taken to ensure that the original meaning of the scene was not altered. In accordance with Cline (1964), the final item selection was done based on a consensus amongst the work group about the consistency of the target responses with the stimulus material and our knowledge about the targets. As a result, items were excluded when the group agreed that an item could not be answered univocally or that the target's responses did not seem to correspond to his or her true mental state as expressed in the task material.

Queried Modalities and Realization of Taxonomic Principles

For each of the eight scenes per scenario, items were designed such that they addressed the target's emotions, cognitions, and the target's relationships to other persons involved in the scene. Sometimes, it was necessary to construct an additional scene if one was not suitable for addressing every type of queried modality. Table 6.3 lists the queried information per modality (i.e., *emotions*, *cognitions*, and *relationships*) and some exemplary item formulations.

Table 6.3

Queried Information per Modality in the Social Understanding Tasks and Example Items

	Emotion		**Cognition**	**Relationship**
Requested information	joy	depression	intentions / goals	dominance*
	anger	compassion	reason	closeness*
	fear	hostility	interest	cooperation*
	sadness	under pressure	importance	activity*
	excitement	happiness	thoughts	familiarity
	disappointment	curiosity	attitudes	sympathy
	contempt	relaxation		typicality
	surprise	disgust		frequency
	indifference	amusement		experience
	concern	grief		
Example item formulation (only ratings-based scales)	"How strongly is the feeling of *anger* present in the target?" from 1 (*not at all*) to 7 (*very strongly*)		*Thoughts:* "How strongly does the target agree with the following thoughts?" from 1 (*not at all*) to 7 (*very strongly*)	*Circumplex:* "How does the target see himself in relation to another person?" from 1 (*very dominant*) over 4 (*neither / nor*) to 7 (*very submissive*)
			Intentions: "How strongly do the following goals determine the target's behavior?" from 1 (*not at all*) to 7 (*very strongly*)	*Others:* "How familiar is the target with this person / this situation?" from 1 (*not at all familiar*) to 7 (*very familiar*)

Note. * Interpersonal Circumplex dimension

This classification finally resulted in a 8 x 2 x 4 x 3 – design, cross-classifying eight targets, two types of settings, four types of task material, and three types of queried modalities. After viewing or listening to the last scene and answering the last questions, participants were asked to rate the personality traits of the target persons on nine dimensions: *(a)* the Big Five personality traits E*xtraversion*, A*greeableness*, C*onscientiousness*, N*euroticism*, and O*penness* and *(b)* the Interpersonal Circumplex dimensions which describe the target's average relationships to his or her social interaction partners on the dimensions D*ominant-Submissive*, A*ctive-Passive*, C*lose-Distanced*, and C*ooperative-Competitive*. The personality ratings cannot be assigned to a special material-related scene but are overall ratings based on the general impression about the targets.

Item and Response Formats

In the first version of the scenarios including only four target persons (Study 1), three different item and response formats were applied, i.e., ratings-based (6-point), multiple choice, and open response formats. The aim of the first main study was to construct reliable and valid scales using various item formats to allow a selection of the most suitable. In the final test version, only ratings-based scales were applied. This decision was mainly based on the empirical results of Study 1 (see Chapter 7.4.2).

Scoring

Target scoring was applied to score the social understanding tasks. This required different scoring procedures dependent on different item formats. The scoring of free response formats was based on a scoring key based on the open answers of the target persons which were assigned a maximum number of credits. Participants' responses were rated in terms of the agreement with the target answers and given full or partial credits for varying degrees of agreement. Two raters were applied which agreed in more 95 % of the cases. Multiple choice items were only applied when the correct answers could directly be derived from the facts collected during material sampling (e.g., the target and another person are siblings). Multiple choice items were dichotomously scored as correct or incorrect (1 / 0).

For the use of ratings-based scales, the original target answer was transformed from the visual analogous rating scale into a Likert-based rating scale, in Study 1 from 1 to 6, and in Study 2, from 1 to 7. This was done by simply assigning a rating category to a target score within a certain range of the analogous scale. Therefore, the 10 cm analogous scale was divided into 100 mm and the target answer received a score from 0 to 100. Consequently, for the 7-point rating scale, a target answer of 0 to 14 received a score of 1, from 15 to 29 a score of 2, etc. For the 6-point rating scale, the calculation was done analogously. To calculate the final accuracy, the difference between the target's and the participant's answer was calculated and assigned a negative sign. Thus, better performance was indicated by a higher score. Finally, the difference derived was weighted according to the maximum possible difference per item. For example, in the 7-point rating scale, when the target answer lay in the middle of the scale (i.e., "4"), the maximum possible difference was "-3" for responses of "1" or "7". When the target answer lay at the extremes of the rating scale, for example "7", a maximum difference of "-6" for a response of "1" was possible. The items were accordingly weighted so that each contributed equally to the total variance of the aggregated scale (e.g., the former item was multiplied by "2" in order to enhance the maximum possible difference from "-3" to "-6").

Some control questions were introduced in Study 2, that succeeded each scenario, which, amongst other objectives, allowed the investigation of the effect of assumed similarity on the performance. The control questions referred to the subjective sympathy of the test takers for, and similarity with, the target person, and to the self-rated cognitive and empathic accuracy (i.e., "How well do you think you could transpose yourself into the target's perspective?" and "How strongly did you identify with the target?"). All control questions were based on a 7-point rating scale.

For the ratings-based scales, group consensus scores were also calculated (i.e., proportion scoring; see Chapter 5.2.4.2) mainly to investigate the effect of this scoring method on the scales' reliability and validity. Possible further scoring options and some additional exploratory research questions will be addressed in the context of Study 2.

6.2.2.2 Social Memory (SM)

Test development for the tasks of *social memory* and *social perception* can be described in a more compressed way. Like stimuli for every task in the SIM, material was sampled in diverse and systematically varied genuine situations involving real persons. The tasks systematically varied the number of involved persons in the stimulus material and the setting (i.e., private vs. public). No further actions were necessary during material sampling such as those that were necessary for the social understanding tasks.

For the tests of *social memory*, material was edited so that sensible and socially relevant entities emerged without changing the original meaning of the scene. The final stimuli had to be capable of providing enough social information to memorize and recall. Most importantly, the later items had to refer only to information objectively present in the stimuli. This implied that no information referring to a person's mental state could be questioned, unless it was presented in written or spoken language (see examples in Table 6.4). In every task, first, the stimuli were presented for a certain time (i.e., text and pictures could be read or looked upon for a certain time; audio or video material could only be watched or listened to once). Directly afterwards, the recall section started. Recall occurred either in free reproduction (for open response formats), by recognition tasks (i.e., for multiple choice formats), or similarly, in a type of paired associates test paradigm. Both, presentation and response times were limited. Answers were scored in terms of the proportion of correct answers for multiple choice items. For open-ended answers, the number of achieved credits in relation to the maximum credits possible was calculated. Comparable to the open answers in the social understanding task, a scoring key provided the rules for the scoring. Table 6.4 presents example items classified according to the underlying task material.

Table 6.4

Exemplary Item Contents Dependent On Task Material

Task material	Example items
Written Language	Based on written correspondence: “What reason does she provide for not writing a postcard during the holidays?” (recognition and reproduction)
Spoken Language	Based on conversations: “According to him, how is he treated by his family?” (recognition and reproduction)
	Based on conversations: memory for previously heard voices (recognition)
Pictures	Based on portrayals of heterosexual couples and colleagues: memory of the correct partner (i.e., paired-associates test paradigm)
	Based on complex picture sequences of interactions: “Which of the following extracts was pictured in the sequence?”, “Who, among the people shown, changed the seats during the sequence?” (recognition and reproduction)
Videos	Based on video clips: “What is the woman’s reaction after the boy has kissed her?”, Which of the following gestures stemmed from the video clip?” (recognition and reproduction)

6.2.2.3 Social Perception (SP)

The social perception tasks also made use of targets. Targets could be persons (i.e., their face or body or their voice), socially relevant written or spoken language material (e.g., an outspoken agreement such as “Yes” or “you’re right”), or more complex social interactions on various communication channels (e.g., the answer to a question in a letter exchange, change of the person who speaks in oral communications, the use of gestures or facial expressions, or interactive elements such as eye contact, the turning to or from someone, common laughter, etc). Comparable to the social memory tasks, only objectively present item contents could be queried. Thus, the target varied depending on the richness of information presented in the stimulus material. Typically, written and spoken material also addressed language content and therefore, information about mental states (directly uttered), which is obvious in Table 6.5 displaying possible targets dependent on task material. Picture- and video-based stimuli can only refer to postures and gestures that stand for certain mental states. Thus, mental states could not be used as targets when pictures or videos were applied.

Table 6.5

Exemplary Targets of Social Perception Dependent on Task Material

Task material	Example items
Written Language	**Target**: "Does the writer criticize a lack of engagement?"; **Stimulus**: "I think you're engagement is great." (true or false?; CRT) **Target**: Has something positive or negative been expressed?; **Stimulus**: "Thanks for you're Email, she lit me up." (positive vs. negative, CRT)
Spoken Language	**Target**: uttered agreement; **Stimulus**: a conversation within which people say, for example, "you're right" or "yes" (unbound reaction, SRT; CRT) **Target**: expression of irony vs. anger; **Stimulus**: real vs. fictional spoken sentences (irony vs. anger, CRT)
Pictures	**Target**: a pictorial portrayal of a person; **Stimulus**: a crowd of people in which the target has to be detected (unbound reaction, SRT) **Target**: eye contact vs. someone is watching someone; **Stimulus**: pictures of small groups of people in interaction (eye contact vs. watching, CRT)
Videos	**Target**: a video portrayal of a person; **Stimulus**: a crowd of people in which the target has to be detected (unbound reaction, SRT) **Target**: turning to or away from the interaction partner; **Stimulus**: videos of two people having a conversation (turning to vs. turning away, CRT)

Note. SRT = simple reaction time task, CRT = choice reaction time task

Prior to stimulus presentation, participants were provided with an example of the target they had to react upon in the upcoming trials. Reactions conventionally occurred by the use of a keystroke or a mouse click. Participants were instructed to react as quickly and accurately as possible. Performance was measured in terms of the time lag between the onset of the target stimulus (i.e., the first appearance of the target within the stimulus trial) and the participant's reaction (i.e., the reaction time, RT). For most tasks, the reaction required a choice between two or three alternatives (i.e., choice reaction time task, CRT) (see Table 6.5 for examples). Some tasks required a free decision of where (indicated by a mouse click on the head of the target) or when (indicated by one keystroke) a target emerged within the stimulus (unbound reaction when only one target was present within one stimulus entity).

Reaction times for wrong responses were excluded from the final score as well as those that were preceded by a false alarm. A false alarm meant that the participant showed a reaction within a short time frame before the eventual target presentation. This could possible bias the eventual response. In addition to the exclusion of these reaction times, some tasks required preparatory data treatment before the final analysis because of skewed distributions or outlier problems. The details of how this treatment was conducted will be presented at the beginning of the results section of each study.

6.2.3 Test Documentation

So far, the test construction principles common to all tasks of one operative ability domain have been described. The scale descriptions, numbers of items, presentation and response times, and concrete item contents, where appropriate, will be included directly in the materials sections of the two main studies. Furthermore, a CD is appended to this work that provides example tasks and the necessary instructions to allow the navigation between the examples. The documentation of the SIM of Study 2 will be present at the Department of Methodology, Diagnostics, and Evaluation Research (Institute of Psychology I) at the Otto-von-Guericke University Magdeburg.

6.2.4 Technical Implementation and Procedures

Technical Implementation

The recordings were done using high-quality technical equipment: a digital video camera (Panasonic NV-GS50), a digital voice recorder (Olympus DM 20, substituted by Sony ICD-ST25 with Stereo Microphone ECM-719), and a digital photo camera (Minolta Dimage A1). The audio stimuli were edited using Cool Edit Pro 2.0, the video files by Pinnacle Studio 9.4. Pictures were finished using Jasc Paint Shop Pro 7. For the final SIM, all tasks were implemented in the Windows® based experimental software Wmc Version 0.18 and the auxiliary software WmcUtil 0.04 based on the programming language Ewx 0.22 which was originally developed for this research project.

In Study 1, only the social perception tasks were implemented in Wmc 0.18 in order to assess the exact reaction times. The other tasks were embedded within the Microsoft PowerPoint presentation software or presented in paper-and-pencil formats. In order to implement all the tasks in Wmc 0.18, which was done in Study 2, some technical specifications were adhered to - all audio files had to be in wave-formats (44 kHz, 16 Bit), videos in mpeg1-formats (384 x 288 pixels) including no audio stream. To designate the exact time of target stimulus onset within a video clip, the respective frame within the video had to be determined. This was done by using the software VirtalDubMod 1.5.10.0. Pictures had to be in bmp-format (maximum 640 x 480 pixels; 8 Bit, 256 colors). These files had to be transformed by using the auxiliary software WmcUtil (see DVD) into the format “sim”. This transformation meant a considerable reduction in the quality of the pictures presented on the full screen. Consequently, the final size of the pictures was reduced to half (320 x 240 pixels) in order to compress the size of the presentation. This also reduced errors in colors and contrasts. All data sampled by Wmc 0.18 were saved as .dat-files that could be imported into SPSS by conventional syntax commands.

All assessments were conducted on PCs with the following system configurations: Windows 2000 professional operating system, 20 GB hard drive, Pentium 4 with 1.7 GHz processor speed, 256 MB RAM, Creative Labs CT4750 sound card, Matrox Millenium G550 graphic card, and 15 inch monitor (60-75 Hz). Additional configurations that were necessary included, *(a)* the deactivation of the antivirus-software in order to avoid regular updates that resulted in the shut-down of the experimental software, *(b)* adjusting the mouse speed to medium, *(c)* adjusting the sound volume to medium, and *(d)* the extended audio settings "spatial" ("Räumlich") had to be switched off. Since the Wmc 0.18 software could not present video files including an audio stream, the self-presentations of the target persons in Study 2 were shown on a large screen, and speech was transmitted over a loudspeaker system (Logitech Z-3 M/N S-0085B; 50-60 Hz). Every PC was equipped with circumaural headphones so that each participant could listen to the auditory material without disturbance.

Procedures

The testing in the pilot and the main studies was done in group sessions consisting of three to nineteen participants. For group sizes larger than 6 participants, two investigators accompanied the testing and gave instructions alternately. In the two main studies, 10-minute breaks were introduced after about one or one-and-a -half hour of testing. General instructions given at the beginning of the first test session introduced the general procedure. Everyone was asked to switch off their cell phones. In addition, subjects received a rule to generate a personal five-digit code in order to allow anonymous testing. The code was also used to combine the data sets from different tasks into the complete data set. Following the general instructions, every task was introduced by the investigators and preceded by example items. This gave the participants the opportunity to ask questions and build up some familiarity with the type of task requirement. For the social understanding tasks in Study 2, one complete example scenario was provided (i.e., target "Birger") involving one exemplary scene per task material. Subjects received monetary compensation for participating in the studies (see Table 6.6). Subjects in Study 2 received half the compensation amount if they requested for detailed feedback about their results. About a quarter of the subjects asked for feedback.

Table 6.6 presents an overview of the studies involving aims concerned with test development and the tasks applied as part of the SIM. The pilot studies took place in August and October 2004 and were particularly directed at the inspection of a new test battery of general auditory tasks (Papenbrock, 2005; Seidel, 2007).

Table 6.6

Pilot Studies, Study 1 and 2: Coverage of the MTMM Design and Aims of Test Development

		Pilot testing August	Study 1 tasks	Study 1 aims	Study 2 tasks	Study 2 aims
SU	V A P F	One scenario (Matthias) Adjustment of test length and type of presentation Formulation of instructions and items	Four scenarios (see Table 6.2)	General test format Item selection based on difficulties and item-total correlations Effect of item formats on scale properties	Eight scenarios (see Table 6.2)[b]	Adding of 4 scenarios, thus, more heterogeneity Item and scale properties Scoring options
SM	V	-	SMv1+2	Item selection based on difficulties and item-total correlations Adjustment of presentation and response times	SMv1+2	Item and scale properties
	A	SMa1[a]	SMa1[b]		Sma1[b] SMa2[a]	
	P	-	SMp1 SMp2		SMp1[a] SMp2[b]	
	F	-	SMf1+2		SMf1+2[b]	
SP	V	-	SPv1[a]	Task implementation General test format Item selection based on difficulties and item-total correlations	SPv1[a] SPv2[a]	Item and scale properties Item selection based on difficulties and item-total correlations for the new tasks
	A	SPa1[a]	SPa1[a]		SPa1[a] SPa2[a]	
	P	-	SPp1[a]		SPp1[a] SPp2[a]	
	F	-	SPf1[a]		SPf1[a] SPf2[a]	
Credit		5 € per hour or course credit	5 € per hour or course credit		Subjects could choose: 5 € per hour or 2.5 € plus detailed feedback of their performance	
Test length SI / total*		1.5 / 4 hours	3.5 / 10 hours		6.5 / 12 hours	
N		29	127		191	

Note. * test length of only social intelligence tasks / total test length including breaks
[a] implemented in Wmc 0.18, including stimuli and answers
[b] implemented in Wmc 0.18, stimuli only; responses in paper-and-pencil format
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

However, the study in August also intended to adjust the test length, modify the formulation of instructions and items, and test the type of stimulus presentation required for the social understanding and social auditory perception tasks. In November and December 2004, the first main study was conducted which assessed four scenarios and tasks of social memory and perception (seven and four tasks, respectively). Thus, not every cell was represented by two tasks in Study 1. This study was specifically directed at investigating the adequacy of item and response formats of the social understanding tasks. Moreover, presentation and response times for the tasks of social memory and social perception should be adjusted. Additionally, the validity of these first versions of the tasks were examined in terms of the structure of social intelligence and the relationship to academic intelligence and personality. Study 2 was conducted between December 2005 and April 2006. The complete SIM was applied including 2 tasks per cell for social memory and perception and eight scenarios. The two written language and video-based tasks for social memory (SMv1+2 and SMf1+2) did not represent different types of tasks in terms of the formal characteristics, as did all tasks in the remaining cells. They only differed in the complexity of task stimuli and were treated as one task in the end. The general objective of Study 2 was to check the effect of task modifications on the scale properties, after Study 1, and replicate the results for validity of the SIM. The study examined whether the newly developed tasks complement the design in a meaningful way. Both main studies, study 1 and 2, required two testing days with five and 6 hours testing time each day.

Data Analysis

All data recorded by Wmc 0.18 were imported by SPSS syntax commands into SPSS. Paper-and-pencil data had to be typed into an SPSS data file by a member of our research project. Data from items that required rating of free responses as per the scoring key were rated by two persons each. In case of disagreement, a consensual rating had to be achieved. This applied, however, to less than 5 % of the responses. Finally, data were analyzed by the use of the statistical software SPSS, SYSTAT, and EQS (Bentler, 1992).

6.3 Research Questions and Hypotheses

Quantitative analyses were not conducted for the pilot studies. They were primarily aimed at finding out about the adequate (technical) implementation in terms of test length, material presentation, and instructions of only a few tasks (see Table 6.6). Thus, the subsequent research questions only referred to the main Studies 1 and 2. Table 6.7 presents an overview of the research questions as they will be outlined in the upcoming passages. The table indicates which study addressed which research question and the applied statistical methods. Additionally, the research questions have been assigned numbers that correspond to those found in the body of the following text and in the headings of the result sections.

Table 6.7

Overview of Research Questions for Study 1 and 2 Including the Statistical Methods for Analysis

Broad aim	**Research question**	**Tasks / Measures**	**S1**	**S2**	**Statistical Methods**
Aim 1: Psycho-metric proper-ties	*1A:* Psychometric properties of items and scales	SU, SM, SP	x	x	Descriptive statistics, reliability analysis
	1A1: Influence of item format on psychometric properties	SU	x		Descriptive statistics, reliability analysis
	1A2: Influence of group consensus scoring on psychometric properties	SU	x	x	Descriptive statistics, reliability analysis
Aim 2: Construct validity	*2A:* Structure of social intelligence	SU, SM, SP			
	2A1: Do the tasks show consistent within-domain correlations? Does data fir to measurement models of the operative ability domains?	SU, SM, SP	x	x	Pearson *r* SEM
	2A2: Does data support the general, structural, and hierarchical models of social intelligence?	SU, SM, SP	x	x	CFA, χ^2-differences test
	2A3: Exploring content-related ability facet	SU, SM, SP		x	CFA
	2B: Convergent construct validity	SU		x	Person *r*
	2C: Divergent construct validity				
	2C1: Do between-domain correlations support the separability of social intelligence from academic intelligence and personality? Does data support separable latent ability factors of social and academic intelligence?	SU, SM, SP AI Personality	x	x	Pearson r, CFA, χ^2-differences test
	2C2: Does social intelligence show structural independency from academic intelligence?	SU, SM, SP AI	x	x	MRA, CFA
	2C3: Does data support a combined facetted model of social and academic intelligence? In which way do the ability facets combine?	SU, SM, SP AI		x	CFA

Table 6.7 *continued*

	3A: Exploring the relationship with self-report questionnaires	SU, SM, SP SR-data	x	x	Pearson *r*
	3B: Exploring the gender differences of social intelligence tasks	SU, SM, SP	x	x	t-test of mean differences
	3B1: Exploring the gender differences for social understanding tasks crossed with the target's gender	SU	x	x	t-test of mean differences, ANOVA
	3C: Exploring scoring alternatives	SU			
	3C1: Exploring the effect of item difficulty on the correlations between group consensus and target scoring	SU		x	Pearson *r*
	3C2: Exploring correlations-based scoring	SU		x	---
	3D: Dimensionality of target score*:* *E*ffect of assumed and real similarity on performance in the SU tasks	SU		x	Pearson *r*
Explora-tory questions	*3E:* Exploring the facetted structure of social understanding	SU			
	3E1: *I*nvestigating the meaning of the ability facets content, modality, and setting	SU		x	CFA
	3E2: Investigating the effect of common target variance	SU		x	CFA
	3E3: Investigating the interaction of task material with the queried modalities on the performance	SU		x	ANOVA
	3F: Exploring the process of social understanding tasks	SU			--
	3F1: Exploring whether performance depended on the point of assessment during one scenario	SU		x	t-test of mean differences, ANOVA, Pearson correlation
	3F2: Exploring the *r*elationship of SU tasks with long term memory of the item material and target persons	SU		x	Pearson *r*

Note. SU = social understanding, SM = social memory, SP = social perception, AI = academic intelligence, SEM = structural equation model, EFA = exploratory factor analysis, CFA = confirmatory factor analysis, MRA = multiple regression analysis, ANOVA = analysis of variance, SR = self-report, S1 = Study 1, S2 = Study 2

It must be noted that Study 1 was first and foremost directed at the test development. Therefore, the emphasis was centered on questions surrounding the psychometric properties and the descriptive statistics of the newly developed tasks. This analysis should allow conclusions about the necessary next steps in test development. The size of the sample and the amplitude and quantity of the tasks applied permitted further analysis surrounding above all the investigation of construct validity. Several further analysis would have been possible. However, it was refrained from investigating any possible research question in Study 1 due to constraints on time and space and due to limitations of data quality.

Ad Aim 1. Psychometric Item and Scale Properties

1A: In line with the principles of classical psychometric test theory, the present work was directed at creating psychometrically sound tasks. This comprised item properties (item difficulty, variance, and item-total correlations r_{it}) and scale properties (reliability indicated by internal consistency assessed by Cronbach's alpha coefficients). It was expected that the internal consistency measures might not be as robust as those found for academic intelligence tasks which typically involve highly homogeneous item contents. The social intelligence tasks were rather heterogeneous. The reliability measures should, however, demonstrate a level of reliability that allows for the aggregation of items into a compound score. The score building of the social understanding tasks in terms of the difference between target's and judge's answers posed another problem. Item difficulty and variance could hardly be interpreted or evaluated since no comparison values were known. Therefore, these parameters could only be evaluated in terms of an internal comparison within the tasks and between Study 1 and 2.

1A1: Study 1 was specifically directed at exploring the effect of different response formats on the psychometric properties of items and scales of social understanding. This analysis was the foundation for the test modifications in Study 2.

1A2: Furthermore, the effect of group consensus scoring on the psychometric properties was examined in Study 1 and 2.

Ad Aim 2. Construct Validity

2A: Structure of Social Intelligence

2A1: The study investigated whether the data supported the structure of social intelligence as specified in the performance model of Weis and Süß (2005). It was expected that the tasks within one ability domain would show coherent correlations with one another. The tasks would load on the respective operative ability factors modeled by structural equation modeling in order to establish measurement models.

2A2: The study hypothesized that a *c*onfirmatory factor analysis provided evidence for the structure of social intelligence. Best data fit was expected for a three-factor model including social understanding, memory, and perception as positively correlated ability factors. Moreover, it was assumed that the tasks would positively load on a general factor of social intelligence. Based on the assumption of a hierarchical model, confirmatory factor analysis would also support a Schmid-Leiman solution and a hierarchical model with a higher-order general factor predicting variance in the second-order ability factors. Figure 6.3

displays the hypothesized models in one diagram. The left side represents the structural model of social intelligence with correlated ability factors. Removing the factor intercorrelations (dashed lines) between the factors and adding the general factor on the right side (dotted lines) results in a hierarchical Schmid-Leiman solution. If possible, the χ^2-differences test would be applied in order to clarify which model provided a better fit to the data .

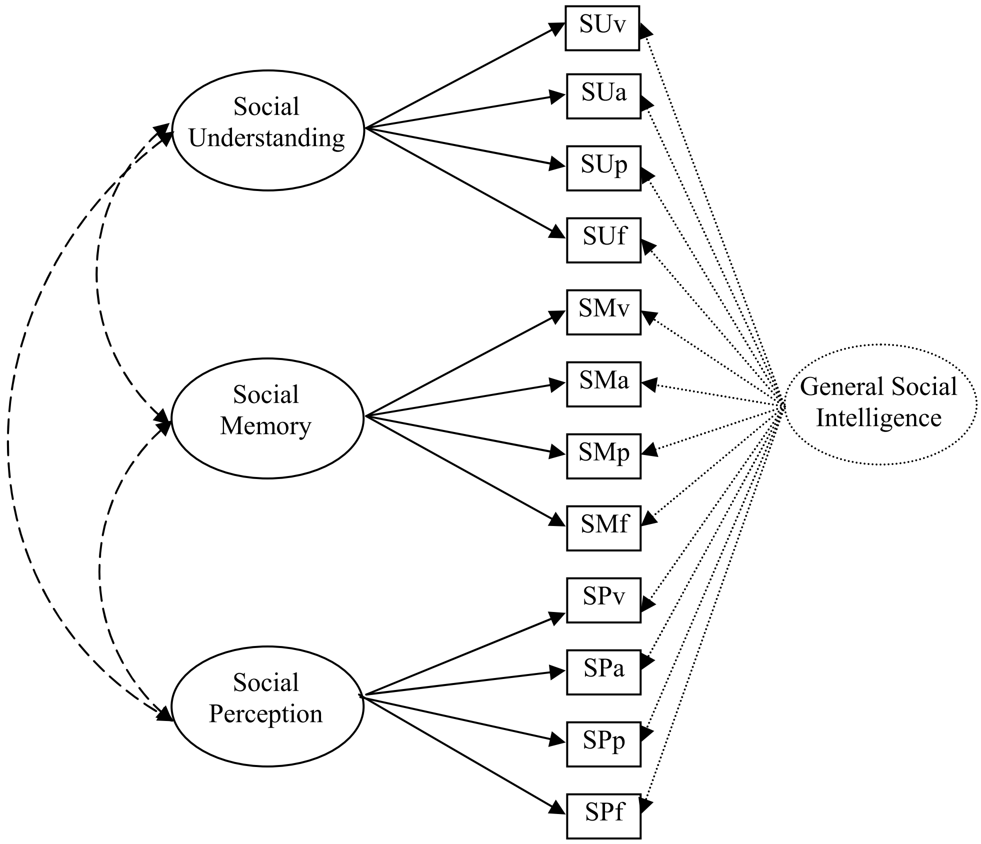


Figure 6.3

Hypothesized Structural Equation Models of the Structure of Social Intelligence

Note. SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos
continuous lines represent the social intelligence measurement models, *dashed lines* indicate the correlations between the ability factors in the structural solution, *dotted lines* indicate elements of the Schmid-Leiman hierarchical solution

2A3: The role of different task material was also of interest. Theory, however, did not suggest whether material-related ability factors could be hypothesized as representing meaningful ability domains. The use of different task material implied different cues or communication channels. Accordingly, it remains an exploratory question whether and how a differentiation of task material resulted in meaningful content-related ability domains (e.g., language-based vs. language-free contents; auditory vs. visual communication channels vs.

semantic contents; etc.). If so, confirmatory factor analysis would support a model with four (or two or three, this number should be explored) content-related ability factors showing loadings of the respective tasks.

2B: Convergent Construct Validity

The relationship of the social intelligence tasks with a measure of nonverbal sensitivity should be investigated. Therefore, the Video Scale of the Profile of Nonverbal Sensitivity (PONS; Rosenthal et al., 1979) was applied in Study 2. Earlier studies, however, could not prove the convergent validity of the PONS itself (Bernieri, 2001; Buck, 1983). Only Bänziger (2005) found a substantial correlation between the PONS and the MERT. Moreover, The Video-PONS showed rather low reliability (see Table 5.3). Thus, positive correlations between the PONS and, particularly, the social understanding tasks could be expected. But a lack of convergent validity evidence would not be seen as problematic for the developed tasks.

2C: Divergent Construct Validity

2C1: The study examined whether social intelligence could be separated from academic intelligence and personality traits. Correlative analysis was expected to demonstrate that within-domain correlations of the social intelligence tasks exceeded across-domain correlations with academic intelligence ability domains (i.e., on a general as well as on specific levels) and with personality traits. It was hypothesized that confirmatory factor analysis would show good data fit for models hypothesizing separate ability factors of social and academic intelligence on different levels of generality (e.g., general academic intelligence and general social intelligence; broad ability domains of academic and social intelligence, etc.) as displayed in Figure 4.6 *(b)*.

2C2: The study also examined if social intelligence would demonstrate structural independency from academic intelligence. Based on a regression-analytic approach, provided that social intelligence could not completely be explained by academic intelligence, some systematic variance had to remain. This could be demonstrated by calculating the identified structural equation models, this time relying on the residuals of the regression analysis when academic intelligence was controlled for. The model was expected to fit the data.

2C3: The study explored whether social intelligence could be classified into an existing model of academic intelligence as a separable and meaningful operative or content-related ability domain. Figure 6.4 presents a model of social and academic intelligence combined in one structural equation model as was postulated in Figure 4.6 *(a)*. According to this diagram, social intelligence was conceived as an additional content-related ability domain within a faceted model of operations and contents. It could be expected that the social intelligence tasks showed loadings on the operative academic ability factors (i.e., Reasoning, Memory, and Perceptual Speed) and on one social content factor so far not included in models of academic intelligence. If this model was confirmed by data, the academic intelligence construct would find a meaningful extension by social intelligence tasks as established in the Structure of Intellect Model of Guilford (1967). This research question was only addressed in Study 2.

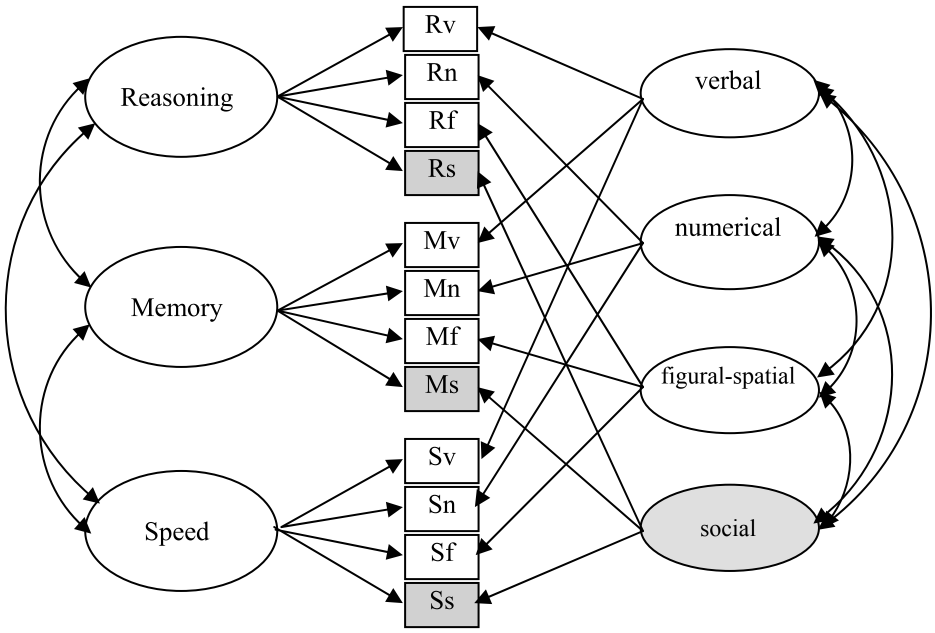


Figure 6.4

Hypothesized Model of Social and Academic Intelligence Integrated in One Faceted Model

Note. white fields represent manifest and latent academic intelligence variables
grey-shaded fields represent manifest and latent social intelligence variables
capital letters: R = Reasoning, M = Memory, S = Speed
small typed letters: v = verbal, n = numerical, f = figural, s = social

3. Further Exploratory Research Questions Not Directly Related to the Aims

3A: Relationship with self-report data

The relationship of the social intelligence performance tasks to self-report questionnaires of social and emotional skills will be explored. Past empirical results, however, do not suggest that substantial correlations will be observed.

3B: Gender differences

The gender differences between social intelligence tasks were examined. Although some authors have interpreted gender differences in favor of women as supportive of the validity of social and emotional abilities (Hall, 2001; Schutte et al., 1998), no hypotheses concerning a gender effect could be formulated.

3B1: Against the background of the already described study of Bronfenbrenner et al. (1958; Chapter 4.3.3.5), gender effects on performance in the social understanding tasks were explored. In particular, the interaction of the targets' and judges' gender was concerned.

3C: Exploring Scoring Alternatives

3C1: Scoring alternatives for the social understanding tasks were explored in order to learn more about the nature of, for example, group consensus scoring or correlations-based scoring in the sense of Snodgrass (2001; see also Colvin and Bundick (2001). The analyses attached to research question *1A2* already addressed the effect of group consensus scoring on the item and scale properties. Additionally, the overlap of target and group consensus scoring was explored. In this context, no hypothesis about the expected overlap was set up because of the conceptual weaknesses of group consensus scoring. It was, however, expected that the overlap was influenced by the difficulty of the items, as could be demonstrated for four fictitious items in Figure 5.3 and 5.4 in Chapter 5.2.4. Thus, it was examined how items of low vs. high difficulty (or items where the modal response in the sample corresponded with the target's answer vs. items where this was not the case) showed a differential correlation pattern.

3C2: Besides group consensus scoring, correlations-based scoring (Snodgrass, 2001) was attempted and its effect on the psychometric properties and the within- and across constructs correlations was examined.

3D: Exploring the Dimensionality of the Target Score

The study examined how much of the target score's variance could be explained by the assumed and the real similarity between judge and target. The indicator of assumed similarity was the similarity rating after each scenario. Real similarity was assessed by a compound score consisting of biographical and personality variables. This analysis was only possible on the level of one scenario.

3E: Exploring the Facetted Structure of Social Understanding

3E1: The scenarios provided a 8 x 2 x 4 x 3 design cross-classifying eight targets, two settings, four material-related contents, and three modalities. Whether the partitioning into four material-related contents could be upheld is left to the empirical results related to research question *2A4* so that changes could occur in this faceted design. Additionally, Study 2 investigated whether the facets of the queried modality and the setting, represented meaningful ability domains. Therefore, confirmatory factor analysis examined the fit of models postulating ability factors related to the facets (i.e., emotions, cognitions, and relationships for the modality facet; private and public for the setting facet). There was reason to assume the factors to be correlated since they were conceptualized as belonging to one social understanding domain. Bernieri (2001), however, accounted for the possibility that factors related to different modalities might be uncorrelated which he did not see as a threat but rather as evidence for a multifaceted social understanding domain.

3E2: The role of common target variance on the identified structure within the social understanding tasks is a methodological question that concerned a possible bias within the tasks of social understanding. The study explored whether meaningful ability domains related to the postulated facets that emerged when variance due to the target persons was controlled for. Structural equation models were set up to postulate eight target-related method-factors and four content-related, three modality-related, or two settings-related ability factors, if these had been supported in the previous analyses. The ability-related factors were hypothesized to be correlated when, therewith, common target variance was bound in the target-related factors. These target-related factors were hypothesized, in contrast, to be uncorrelated

3E3: Archer and Akert (1980) and Ekman et al. (1980) had investigated the effect of the availability of cues on subjects' performances in tasks of social abilities. In accordance with their research question, the current study examined whether items assessing different modalities showed different degrees of difficulty depending on different underlying task material. It was possible that different task material (different communication channels) only allow specific judgments related to specific modalities (e.g., a judgment about the thoughts of a target might only be possible when language-contents were provided in written or spoken language). However, the scope of this analysis was surely restricted because different items were underlying the different modalities and contents so that the result could as well be an effect of the item characteristics.

3F: Process During Social Understanding Tasks

3F1: The study examined whether there was an increase in the accuracy of judgments during the course of one scenario task. This could be interpreted as accumulated knowledge about the target, getting broader and deeper with every new stimulus. If an effect of the course within one scenario on the accuracy could be observed this could be attributed to the presence of an underlying knowledge structure.

3F2: The study investigated whether performance in the social understanding tasks was related to long term memory of the task material and the background information about the targets in the social understanding tasks. If higher long term memory of only the task material in the social understanding tasks was associated with better performance in the scenarios, this would carefully point towards perception-based bottom-up processing as a meaningful determinant of social understanding abilities (Bless et al., 2004; Buck, 1983). If performance was related to long term memory of the background information, this would suggest knowledge-based top-down processing (Bless et al., 2004; Buck, 1983).

7 Study 1

7.1 Sample

One hundred and twenty seven students of the Otto-von-Guericke-University Magdeburg in Germany participated in the present study. One hundred and twenty four had graduated from German high school after the 13th grade (i.e., “Gymnasium”) as per regular schedule, one had graduated early with a “Fachhochschulreife” after the 12th grade, two did not provide information about their educational status. Participants’ areas of study at university were mainly first-year psychology, economic sciences, and mathematics. They were recruited in university classes and received monetary compensation (see Table 6.6) or course credits. One participant did not return after the first day of testing, this half data set was excluded from further analysis. The mean age of the participants was $m = 21.35$ ($sd = 3.06$), and 53.5 % (i.e., 68) of the participants were females.

7.2 Material

7.2.1 Social Intelligence Tasks

The general ideas and principles underlying the construction of the tasks have already been described and will be repeated in the upcoming task descriptions. During the process of test development, the responsibilities for test construction were partitioned between the present author and Kristin Seidel (Seidel, 2007). Kristin Seidel was responsible for the construction and description of the tasks based on written and spoken language. A detailed description of the tasks and the considerations underlying the construction and modifications can be found in Seidel (2007). The partitioning was, however, not possible for the social understanding tasks because of the scenario approach. These tasks will therefore be addressed in both works.

The following passages present the newly developed social intelligence tasks and the assignment to the cells of the MTMM design. The assignments simultaneously represent the abbreviations used to refer to the respective tasks in the Tables and the results sections. A list of abbreviations is included in the index of abbreviations.

Social Understanding Tasks

The scenario approach has already been described in detail in a previous section (see Chapter 6.2.2.1) and only the core features will be outlined at this point. In short, the scenario tasks required subjects to judge the emotions, cognitions, relationships, and the personality of target persons on a 6-point rating scale. To operationalize every content domain, judgments were based on information from written and spoken language, pictures, and videos. In the first study, four scenarios were applied presenting Renate (*SU_RF*), Christoph (*SU_CP*), Katharina (*SU_KL*), and Matthias (*SU_MM*) as target persons (see Table 6.2). Three different item formats were used (i.e., free response, multiple choice, and 6-point Likert-based rating scales). Material-related task contents were systematically varied within and across one scenario. It turned out to be difficult, however, to vary the item formats and the queried modality within and across the scenarios because some modalities were favored by a certain item format (e.g., the cognitive modality was better represented by free response formats, and emotions were better queried by the use of ratings-based scales). Therefore, the following cross-table (Table 7.1) emerged showing the number of items across the four scenarios classified in terms of the task material, the queried modality and the item formats.

Table 7.1

Number of Items in Social Understanding Tasks Classified into Contents, Modalities, and Item Formats (Study 1)

Queried modality	Item format	SUv	SUa	SUp	SUf	∑	∑
	Rating scales	31	38	8	24	101	
Emotion	Multiple choice	0	0	0	0	0	105
	Free response	0	2	0	2	4	
	Rating scales	2	2	2	0	6	
Cognition	Multiple choice	0	0	0	0	0	20
	Free response	2	8	2	2	14	
	Rating scales	18	16	5	46	85	
Relationship	Multiple choice	9	1	9	11	30	122
	Free response	2	1	1	1	7	
	∑	64	68	27	86		

Note. SU = social understanding, v = written language, a = spoken language, = pictures, f = videos

Table 7.1 also presents the common abbreviations used hereafter for the newly constructed scales (see Note to Table 7.1). The scenarios contained between 27 and 86 items related to different task material. At the end of each scenario, the personality traits of the targets had to be rated on a 5-point rating scale on nine dimensions (i.e., the Big Five and four of the Interpersonal Problems – Circumplex). Performance was scored by target scoring (i.e., the weighted difference from the target answer). The average duration for each scenario was estimated to be 20 minutes. Data sampling, however, showed large variation between the testing times of different subjects. At times, a scenario lasted about half an hour. This was, in part, due to some subjects who watched, listened to, or read the scenes more than once which was possible in the PowerPoint presentation format. The final reliabilities (Cronbach's alpha) of the scales were .75 / .68 / .65 / .76 for the written and spoken language, pictorial, and video-based scales, respectively. How these scales were eventually composed will be presented when the psychometric properties of the scales are described (see Chapter 7.4.2.1).

Social Memory Tasks

Social Memory – written language (SMv1+2): Memory for Written Correspondence

Participants were presented four written one-way or two-way correspondences (e.g., a letter written after a skiing vacation) and told to memorize as many socially relevant details as possible. Reading times varied between 1:30 min. and 3:20 min. Participants were then asked to freely reproduce the information from the texts in free response format items (e.g., "What does the writer say that she felt sorry for?"). Response time was limited (between 2:00 min. and 2:30 min.). The test contained a total of 38 items and was subdivided into two halves for data sampling. The task showed a reliability of .74 (Cronbach's alpha) in the present sample.

Social Memory – spoken language 1 (SMa1): Memory for Conversations

Participants listened to twelve audio-recorded monologues and conversations between two or more people (e.g., a male person talking about a conflict at work with both colleagues and superiors) and were told to memorize as many socially relevant details as possible. The length of the stimuli varied from 1:36 min. to 2:14 min. Participants were then asked to freely reproduce the information from the recordings in free response format items (e.g., "According to the speaker, how did the superior react?"). Response time was limited to 1:00 min. The test contained 62 items. This test was also subdivided into two halves for data sampling. Cronbach's alpha for the present sample was .78.

Social Memory – pictorial 1 (SMp1): Memory for Couples (Weis & Süß, 2007)

This task required subjects to observe and memorize pairs of people as well as possible. Subsequently, participants were shown one of a pair and were asked to identify the correct partner from four alternatives. All persons displayed wore different clothing in the recall section. This task was applied in the study reported in Weis and Süß (2007). Te pictures were presented in blocks. In contrast to three blocks in the original version, two blocks of eight pictures each were presented in the present study. Weis and Süß (2007) had only included pictures of heterosexual couples. In the present modifications, only the first block portrayed heterosexual couples. In the second block, pictures showed pairs of colleagues of the same gender (i.e., four pictures with male, four with female colleagues). The recall section followed directly after one presentation block. Every picture, including the presentation and the recall sections, stayed on the screen for three seconds. Subjects' response time was also limited to three seconds. Subjects indicated their answer by marking the correct alternative on the answering sheet. Performance was scored as the proportion of correct answers. The reliability of the task in the present sample was .52.

Social Memory – Pictorial 2 (SMp2): Memory for Situations

This task was a first version applied to test the item and the general test format. The total testing time available was restricted to a few minutes. Therefore, consecutive to the task SMp1, a first short version of the present task was administrated. Participants were required to memorize as many socially relevant details out of a sequence of pictures showing actions of different numbers of people in one context (e.g., the first sequence showed three pictures of family members at a family get-together having dinner and taking a boat trip). Pictures stayed on the screen for five seconds each. The first sequence of the family get-together involved five different persons. Pictures of the second sequence showed teachers during their break carrying out different activities in the staff room (e.g., talking to each other, having lunch, changing seats, etc.). This sequence consisted of nine pictures involving a total of 16 people. After viewing the sequence, participants had to answer open-ended and a few multiple choice questions about socially relevant details. Figure 7.1 presents one example picture extracted from the second sequence and an associated example item. The first sequence comprised seven items, and the second sequence eight items. Answering time was limited to 1:00 and 2:00 min. respectively, for sequence 1 and 2. The testing procedure showed that participants had problems answering the questions within the given time frame. The free responses were rated by two raters in terms of the accordance with the correct answer. Performance was scored in terms of the number of achieved points in relation to the maximum number of points. Multiple choice answers were scored as proportion correct. Cronbach's alpha in the present sample was .14.

Within a sequence of nine pictures, subjects also see the following one.

Example item:

"Which of the three women sitting side by side were talking to each other?" (open-ended response)

Figure 7.1

Example Picture and Item for Social Memory – pictorial 2 (SMp2) (Study 1)

Social Memory – video-based 1+2 (SMf1+2; Weis & Süß, 2007): Memory for Situations - Videos

Only one type of task assessed social memory based on video material. This test approach was presented in the study conducted by Weis and Süß (2007). Two levels of complexity could be differentiated within the task associated with the number of people involved in the video scenes (i.e., less than five people representing SMf1 and more than four people SMf2). No separated analyses, however, were conducted for the two sub-tasks. The task comprised of four video scenes presenting different social events (i.e., a dinner, an excursion of a kindergarden group, an exhibition, and a bowling night). Participants viewed each video scene once and were instructed to memorize as many socially relevant details as possible. After viewing each scene, subjects had to answer mostly open-ended and some multiple choice questions. Table 7.2 presents an overview of the four scenes including example items and the number of items. Answering time was limited to 1:15 min. for each scene. Testing, however, showed that the answering time was too short to allow answers to every item. The free responses were rated by two raters in terms of the accordance with the correct answer. Performance was scored in terms of the number of achieved points in relation to the number of available points. Multiple choice items were scored as proportion correct. The reliability of the task in the present sample was .47.

Table 7.2

Video Scenes in the Social Memory – Videos Task, Example Items and Video Length (Study 1)

Content	Scene description	Example item	Video length	Item count
1: Dinner	A family, consisting of parents and an elder sister and a younger brother, is having dinner during their vacations in Spain. They are eating Paella.	"Who did the young boy kiss during the dinner?" (free response format)	2:02	9
2: Kindergarden excursion	A group consisting of several adults and children is seen at a barbecue.	"How many children have their barbecue sitting together at one table?" (free response format)	2:39	7
3: Exhibition	Three people are observed talking to each other during the welcome speech of an exhibition.	"Do the woman and the man on her right side look into each others' eyes?" (multiple choice)	1:38	7
4: Bowling event	A group of seven adults is seen sitting together at one table at a bowling event.	"How does the woman returning to the group learn what has happened while she was away?" (multiple choice)	1:48	9

Across the social memory tasks, the following taxonomic classifications were achieved (see Table 7.3). Some problems concerning the accomplishment of the taxonomic principles and the accounting for socially relevant item contents were encountered during test construction. These problems were associated each other. Private and public settings could be enclosed in every single task. One task (SMp1) included situations with dyads only. Two tasks (SMp2 and SMf1+2) involved groups of people. Thus, a lower level of complexity (i.e., task material presenting only a single person) was not achieved since it turned out difficult to include socially relevant item contents when only one person was displayed. In this respect, just querying the physical appearance of the person or single movements did not seem to be socially relevant. The focus was therefore the stimuli containing small groups and pairs of people. Another problem regarded the transformation of pictorial and video-based material into items and responses based on written language. It turned out to be difficult to formulate items so that concrete pictorial stimuli contents were addressed univocally. For example, one scene included a male person leaning forward onto a table, the related item asked for the posture of this man and subjects were to describe the physically present posture. However, several responses included descriptions of how the man appears while leaning onto the table (e.g., relaxed, interested, etc.). Moreover, subjects' responses sometimes showed a large range of possible interpretations because of the same type of problem.

Table 7.3

Taxonomic Principles Underlying the Social Memory Tasks Based on Pictorial and Video Material (Study 1)

		Setting	
		Private	**Public**
Number of persons involved	**Two people**	SMp1 block 1: heterosexual couples	SMp1 block 2: colleagues
	Small groups (number of persons in parentheses)	SMp2 sequence 1: family excursion (5) SMf scene 1: dinner (4) SMf scene 4: bowling night (6)	SMf scene 3: exhibition (3) SMp2 sequence 2: teacher (16) SMf scene 2: kindergarden excursion (16)

Note. SM = social memory, p = pictures, f = videos

Social Perception Tasks

Social Perception – written language (SPv1): Perception of Social Cues in Texts

Subjects were presented one or two written target statements or questions on one half of the screen (e.g., “Does the sender criticize a lack of engagement?”). Subsequently, a short text was presented on the other half of the scree two seconds after the target statement (e.g., “I think you’re engagement is great.”). Based on this text, subjects had to make a decision about the truthfulness of the statements (“true” vs. “false”) or whether the question could be answered with “yes” or “no” (i.e., choice reaction time), as quickly as possible. For the case of two presented target statements or questions, both needed to be true for the answer to be “true”. Target statements or questions and the text could be seen simultaneously for a limited time. Subjects indicated their decision by pressing the respective key representing their choice (e.g., “<” for false, “-“ for true). Response time was limited to 12 sec. Performance was scored in terms of the reaction time for correct responses accounting for false alarms. The task consisted of 35 items presented in one row without breaks. Cronbach’s alpha reliability coefficient for the present sample was .83.

Social Perception – auditory (SPa1): Perception of Social Cues in Spoken Language

Subjects were made to listen to extracts from audio-recorded conversations (e.g., a telephone conversation between two male friends about study concerns). Prior to the presentation of the recordings, they were instructed to attend to specific target cues. Target cues could vary according to the complexity (e.g., a laughter, mention of a given name, interruptions, or agreement, etc.). Subjects had to react as quickly as possible as soon as they perceived the respective cue within the recordings. Either one or more than one cue had to be attended to (simple vs. choice reaction time). Subjects indicated their decision by pressing the respective key representing their choice (e.g., “<” for mentioning a name, “-“ for agreement). Answering time was limited in terms of the time until the next cue emerged within the recordings. Performance was scored in terms of the reaction time for correct responses accounting for false alarms. The task consisted of 13 audio recordings and a total of 136 items. Cronbach’s alpha in the present sample was .88.

Social Perception – pictorial (SPp1): Person Perception – Pictures

Subjects had to detect given target persons within pictures of crowds at different public locations (e.g., a market place, a store, a mall, a pedestrian zone, etc.). Prior to the test trials, the targets were presented with portrayals showing the persons’ whole body. Afterwards, subjects had to indicate as quickly as possible the location of the target within the crowd by a mouse click on the target’s head. Targets wore different clothing in their presentation and in the trials. The position of the targets within the pictures varied unsystematically. The task comprised three practice blocks and seven test blocks, each block contained ten items. Either one, two or three targets had to be attended to within one block. One picture, however, displayed only one target person. Answering time was limited to 10 seconds. Performance was scored in terms of the mean reaction time of correct trials. Cronbach’s alpha in the present sample was .90.

Social Perception – video-based (SPf1): Person Perception – Videos

Subjects had to detect target persons within videos of crowds at public locations (e.g., a market place, a store, a mall, a pedestrian zone, etc.). The item presentation was preceded by videos of the targets showing their whole body. Afterwards, subjects watched videos of crowds and had to react as quickly as possible when they detected the target. Either one, two or three targets had to be attended to within one block (simple or choice reaction time). Targets wore different clothing in their presentation and in the trials. Subjects indicated their response by pressing the respective key (e.g., “<” when target person 1 appeared in the video, “-“when target person 2 appeared in the video). The task comprised one practice block and three test blocks, containing ten items each. Answering time was limited in terms of the video length (i.e., no reaction was possible after the video had stopped). The length varied between 8 and 47 seconds. The target appeared in the videos at varying points in time. The point of emergence was determined in terms of the respective frame in the video. For example, a video of 10.96 seconds consisted of 274 frames, one frame representing 40 ms. Target appearance across all videos ranged from the first frame to frame 667 so that subjects did not build up expectations about the target appearance. Performance was scored in terms of the mean reaction time of correct trials accounting for false alarms prior to the target appearance. Figure 7.2 presents an example video stream showing the start and the end of a video and the point of target appearance. Reactions prior to that point were scored as a false alarm (e.g., a possible distracter person prior to target appearance). Valid reactions could only occur after the target appeared. The reliability coefficient in the present sample was .70.

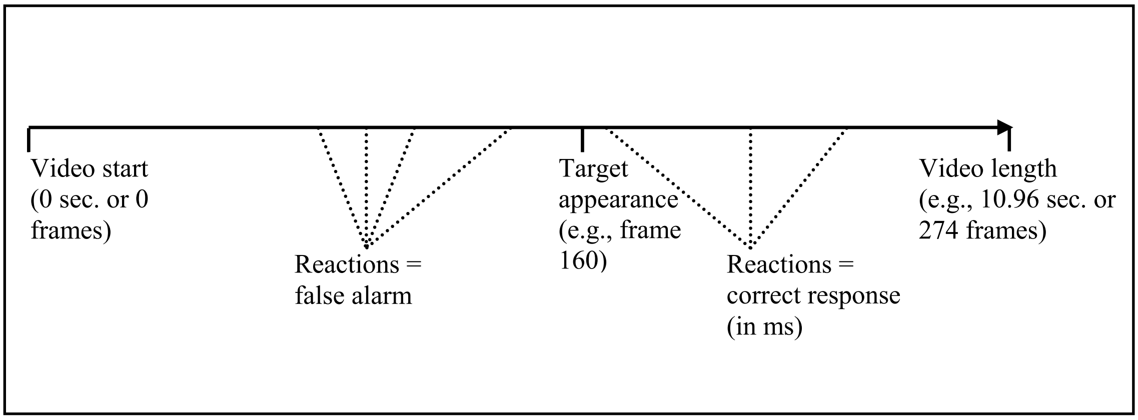


Figure 7.2

Schema of Scoring Reaction Times and Accounting for False Alarms of the Task SPf1 (Study 1)

7.2.2 Validation Instruments

Academic Intelligence: BIS-Test (Jäger et al., 1997)

The BIS-Test was assessed as a test for academic intelligence. The test was based on the BIS-Model which represents a faceted hierarchical model cross-classifying an operational and a content facet with seven broad ability factors (i.e., operational facet: BIS-R = *Reasoning*, BIS-M = *Memory*; BIS-S = *Speed*, and BIS-C = *Creativity*; content facet: BIS-V = *verbal ability*, BIS-N = *numerical ability*, and BIS-F = *figural-spatial ability*) and resulting twelve cells. The BIS-Model has been described in Chapter 4.1. The structure of the BIS-Model assessed by the BIS-Test has been replicated several times (Beauducel & Kersting, 2002; Brunner & Süß, 2005; Süß et al., 2002). The manual reported high internal consistency coefficients. Test-retest reliability coefficients ranged from .65 for BIS-C to .90 for BIS-R (Jäger et al., 1997).

The complete BIS-Test comprised of 45 tasks assigned to the twelve cells, each including requirements of one operational and one content domain. In Study 1, however, the complete test was not applied due to time constraints and due to an arrangement with a diploma student (Feigenspan, 2005). Her thesis concerned the development of a computer-administrated BIS-Test version. She partly relied on the present sample to examine the equivalence of the computer-administrated and the paper-and-pencil version. Therefore, tasks that could not be administrated completely on the computer were not applied. Tasks for BIS-C (twelve tasks) and one figural-spatial task from a total of 15 tasks of BIS-R were omitted because they required open response format. Likewise, some tasks for BIS-M and BIS-S were excluded. Finally, BIS-M was assessed by seven tasks (two, two, and three, for the verbal, numerical, and the figural-spatial cell), BIS-S was also assessed by seven tasks (two, three, and two, for the verbal, numerical, and the figural-spatial cell). A total of 28 tasks and one warming-up task were assigned to five parts that were administrated at different points in time throughout the two testing days (see Appendix C for the order of testing including instruction and working time per task). The five parts lasted about 15 minutes each and included between five and six tasks. The reliability coefficients (Cronbach's alpha) for the six broad ability factors in the present sample were .72 / .70 / .81 / .65 / .73 / .82 for BIS-S / -M / -R / -F / -V / -N, respectively.

Personality: NEO-FFI (Borkenau & Ostendorf, 1993)

The NEO-FFI was applied as a self-report measure of the Big Five personality traits *Neuroticism (NEO-N)*, *Agreeableness (NEO-A)*, *Extraversion (NEO-E)*, *Conscientiousness (NEO-C)*, and *Openness (NEO-O)*. The questionnaire consisted of 60 items. Subjects had to rate their agreement with a given statement on a 5-point rating scale (from 0 "*I completely disagree*" to 4 "*I completely agree*"). The final scales consisted of the mean ratings over all items per scale accounting for the number of missings. Reliability coefficients reported in the manual (Cronbach's alpha) were .85, .80, .71, .71, and .85 for (NEO-N, -E, -O, -A, -C, respectively). Results from the present study basically supported these findings and showed alpha coefficients of .81 (NEO-N), .76 (NEO-E), .72 (BIS-O), .72 (NEO-A), and .86 (NEO-C).

7.2.3 Instruments Peripherally Related to Research Questions

Baseline Measures

Simple Reaction Time Task (SRT; Oberauer, Süß, Wilhelm, & Wittmann, 2003; Sander, 2005)

The SRT task version applied in the present study was extracted from the Working Memory test battery presented in Sander (2005). The task was implemented in the experimental software Wmc 0.18. Stimuli were presented and reactions sampled only by the use of the computer. Subjects had to react as quickly as possible with a keystroke on the space bar as soon as a white dot appeared on the screen. Prior to the dot, a small white fixation cross emerged to direct the subjects' attention to the upcoming stimulus. The time delay between the presentation of the fixation cross and the dot varied between 970 and 2970 ms. The task consisted of five example trials and ten blocks including five test trials each. Subjects performance was scored as the mean reaction time for accurately accomplished trials. A trial was accomplished accurately when a reaction occurred within a certain time frame after a dot presentation and without a prior false alarm. Cronbach's alpha coefficient in the present study was .96. The SRT was intended to serve as a baseline measure for the social perception tasks in which keystrokes were required as responses.

Mouse Speed Task (MT; Oberauer et al., 2003; Sander, 2005)

The Mouse Speed task also stemmed from the Working Memory test battery presented in Sander (2005). The task assessed the subjects' baseline speed in moving the mouse pointer from a preset position to an occurring dot on the screen. Prior to the dot, a fixation cross announced the upcoming stimulus. The position of the dot could vary over the whole screen. The task was implemented in the experimental software Wmc 0.18 and administrated on the computer only. The task consisted of ten example trials and three blocks of 25 trials each. Performance was scored as the mean reaction time of correct trials. A correct response required hitting the dot with the mouse pointer within a range slightly larger than the dot boundaries and no early or late reaction outside of a predefined reaction frame. The Mouse Speed task showed high reliability coefficients (Cronbach's alpha) of .98 for the present sample. The task served as a baseline measure for the social perception tasks that required a mouse reaction (SPp1).

Readspeed (Rüsseler & Münte, 2001)

The Readspeed task was applied as a baseline measure for the social perception tasks requiring the quick reading of words or sentences (SPv1). It was extracted from research on reading disabilities. The task consisted of a meaningful text containing 198 words presented one after another. Subjects had to press the space bar as soon as they had read the word. After the keystroke, the word disappeared and the next one emerged. Subsequent to the text, three multiple-choice items about the text contents controlled whether the subjects had pressed the space bar without reading the words. Performance was scored in terms of the mean reading time for one word. Cronbach’s alpha reliability coefficient in the present sample was .996.

Self-report Questionnaires

Biographical Questionnaire

Biographical data for each participant were sampled by a questionnaire containing items that addressed the age, gender, level of education, and the high school grades per subject-matter reported in the leaving certificate. Additionally, the general musical experience was dichotomously scored (“Do you have musical experience? Yes or No?”) and a self-perceived hearing capability was assessed (dichotomous item: normal vs. disabled hearing capability).

Computer Experience (Feigenspan, 2005; Süß, 1996)

The items were adapted from questionnaires applied in previous studies (Feigenspan, 2005; Süß, 1996). They referred to the individual experiences in dealing with a computer in: *(a)* the time period for which the subjects were now familiar with computers (in years), *(b)* the average time spent at the computer (multiple choice: several hours per day, week, month, or less), *(c)* the life-context of using the computer (e.g., work-related or private activities), *(d)* the self-assessed amount of knowledge and skills in using the computer for different purposes (i.e., data and text processing, internet and email services, programming, course-related activities, etc.), and *(e)* the self-assessed experience in relation to people of the same age group. Computer experience was conceived as a possible performance determinant of computer-administrated tasks. It was particularly relevant for the study by Feigenspan (2005) which tested a computer version of the BIS.

Social Behavior Questionnaire (Amelang et al., 1989)

The Social Behavior questionnaire aimed at assessing self-reported socially intelligent social behavior, operationalized by 40 prototypical behavioral acts. These were derived from a study based on the implicit theory approach (Amelang et al., 1989; see Chapter 5.1.1) and represented prototypical socially intelligent actions. Subjects rated each act in terms of how often they had performed this act themselves in the past on a 4-point rating scale from 1 ("*never*") to 4 ("*frequently*"). The compound score consisted of the mean ratings over all items accounting for the number of missing values. The reliability coefficient (Cronbach's alpha) for the present study was .83.

7.2.4 Instruments Not Related to Research Questions

Several instruments were applied during data sampling that were not related to any research question or analysis in the present study. They will not be described in detail but are listed below, and include information about the intended measurement constructs. More information for most of the tasks is available in the doctoral thesis of Seidel (2007) which provides details about the tasks and the associated analyses.

Questionnaires

- *Masculinity-Femininity* (Hathaway, McKinley, & Engel, 2000): a 56-item self-report about the tendency to exhibit typically male or female interests or behavior.
- *Hearing Screening Inventory* (Coren & Hakstian, 1992): a twelve-item self-report about the subjects' hearing sensitivity in everyday life.

General Auditory Tasks – Nonverbal Tonal Tasks

- *Recognition of Repeated Tones* (Stankov & Horn, 1980): a 17-item performance task to assess the ability to recognize one tone within a sequence of eight tones that is played once.
- *Tonal Analogies* (Stankov & Horn, 1980): a 17-item performance task that measures the ability to identify a pitch difference, from among 4 alternatives, that was equivalent to a previously played pitch difference.
- *Tonal Figures* (Stankov & Horn, 1980): a 17-item performance task to assess the ability to recognize accordance in the tonal composition of tonal sequences between a target sequence and four alternative sequences.
- *Tonal Series* (Stankov & Horn, 1980): a 21-item performance task that assessed the ability to identify the logical completion of a tonal series of four tones by selecting the correct tone from four alternatives.

- *Chord Decomposition* (Stankov & Horn, 1980): a 14-item performance task measuring the ability to identify the correct three single tones that constitute a previously played target three-tone-chord.

General Auditory Tasks – Nonverbal Tasks

- *Auditory Inspection Time Task – Pitch / Loudness* (Deary, Head, & Egan, 1989): In general, inspection time tasks assess the discrimination ability between two stimuli only differing to a predetermined varying degree in one attribute (e.g., two tones only differ in their loudness between 50dB and 55 dB). The present study included two auditory inspection time tasks. One required the discrimination of pitch levels (AIT-P), and the other of levels of loudness (AIT-L), containing 120 items.
- *Rhythm Reproduction* (Stankov & Horn, 1980): a 20-item performance task that required participants to reproduce rhythms by a keystroke on one key. Rhythms could vary in length and complexity.
- *Sound Recognition*: a 20-item performance task requiring the recognition of 20 previously heard target sounds out of 45 sounds.

General Auditory Tasks – Language-Based Tasks

- *Recognition of Repeated Voices* (Stankov & Horn, 1980) : a 25-item performance tasks that assessed the ability to recognize the one voice within a sequence of eight voices (speaking the same word) which was heard only once.
- *Masked Words* (Stankov & Horn, 1980): a 35-item performance task assessing the ability to identify spoken words (e.g., “tree”, “table”, etc.) against a noisy background varying in intensity (i..e., typical noise of a party), responses had to be written down in free format.
- *Audiobook*: a 14-item performance task requiring subjects to reproduce as many details as possible from a previously memorized text, the text was played twice and contained facts about a specific topic (i.e., a report about a journey to Macao) without including any socially relevant details; answering time was limited to 3 minutes.
- *Dissected Sentences* (Stankov & Horn, 1980): a 19-item performance task assessing the ability to rearrange previously disarranged words in order to compose a meaningful sentence. Participants had to freely produce the sentences and write them down.

7.3 Procedures

The procedures common to both studies have been described in Chapter 6.2.4. Appendix C presents the order of administration and the planned duration of the single tasks for testing days 1 and 2. The order involved changing requirements, concerning different task contents (e.g., auditory vs. video-based tasks) and operations (e.g., reasoning requirements vs. self-reports, etc.). Baseline testing took place prior to the tasks that the baseline measurement was directed at. Testing per day was planned to last about five hours and was partitioned into four sessions including three breaks of about ten minutes each. Some tasks, particularly the scenario tasks, however, showed large variance in testing time due to single participants (see the task description). Since each task began and provided instructions to the entire group at the same time, testing time increased by an average of half an hour per day.

7.4 Results

7.4.1 Preparatory Data Analysis

Before starting the main analysis of the present study, data of the newly constructed tasks were screened for missing values and distributions. Fidell and Tabachnik (2003) have stressed the importance of such preparatory data screening processes to exclude violations of assumptions for applying certain statistical analyses.

1. As a first step in data screening, items and subjects were excluded when the number of missing values exceeded 15 % of the data points. Additionally, outlying data points were inspected and treated only if they influenced bivariate distributions or if an implausible deviation from the sample mean occurred.

The following steps were taken for tasks based on *reaction time scores* (i.e., only social perception measures and baseline speed measures):

2. Trials based on a wrong answer were set to missing.
3. Trials preceded by a false alarm were set to missing.
4. Reaction times lower than 100 ms were set to missing.
5. If necessary, based on a screening of the distribution of a first compound score, reaction times of single trials slower than 3 *sd* above the mean of the sample were set equal to 3 *sd* (trimming of outliers on the group level).

6. If necessary, reaction times of single trials of 3 *sd* above the mean of the individual were set equal to 3 *sd* above the individual mean (trimming on an individual level).

The following paragraphs present the results of the data screening and, if necessary, the results of any preparatory data treatment. Analogous to the test construction, the present work focused on the social understanding tasks and on the pictorial and video-based tasks of social memory and perception. For a detailed description of the scale construction and data treatment of the tasks based on written and spoken language, see Seidel (2007).

Social Understanding Tasks

The number of missing values in the social understanding tasks ranged from 0 to 16 per item for 126 participants which was less than 15 % of the data points per item. However, four subjects exceeded the threshold of 15 % missing values (between 44 and 94 missings for 259 items). Since these were newly constructed tasks without any possibility to refer to established empirical results, the subjects were excluded from further analysis.

Social Memory – SMp1

No actions were required because just one data point was missing in the Memory for Couples task SMp1.

Social Memory – SMp2

As expected from the experiences during data sampling, the task Memory for Social Situations – pictures (SMp2) showed large numbers of missings. Out of 15 items, just four items had less than 15 % missing values (nine items had more than 50 % missing values). Therefore, only the four items were used for further analysis. After excluding the items, another 27 subjects still had more than 15 % missing values. This amounted to more than one fifth of the total sample. Therefore, this task will be omitted from any consecutive multivariate analysis because of a significantly large reduction of the sample size. However, the psychometric properties of the task will be reported hereafter.

Social Memory – SMf1+2

Comparable but less severe problems occurred in the Memory for Social Situations task based on videos (SMf1+2). Data sampling had shown that subjects had difficulties working on every item because of short presentation and response times. Of the original 31 items, 16 had more than 15 % missing values. In the remaining 15 items, eleven subjects had more than 15 % missings and were excluded from further analysis. Again, no substitution of missing values was performed because of the lack of experience with the task.

Social Perception – SPp1

The number of missing values in both tasks of social perception (SPp1 and SPf1) was counted before any recoding or transformation was executed according to steps 2 through 6 explained above. Thus, a missing value represented an item that had not been worked on (i.e., absence of a response during predefined answering time). Of the original 100 items on the pictorial task, only 4 had more than 15 % missing values and were excluded. No subjects showed more than 15 % missing values. According to step 2, the reaction times of wrong responses were set to missing. As expected, the task was easy to accomplish. Figure 7.3 shows the distribution of the number of wrong responses (i.e., a mouse click not on the head of the target person). About 75 % of the subjects had six or less missing values because of wrong answers, and 50 % of the subjects had less than four.

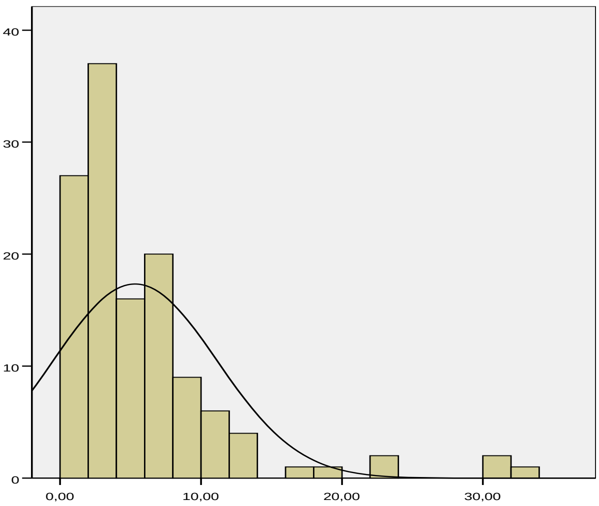


Figure 7.3

Histogram of Missing Values Due to Wrong Responses in SPp1 (Study 1)

Before composing the final score, items with wrong responses more than the threshold value were excluded (i.e., this applied to ten items). A first screening of the distribution of the compound score of all remaining items showed a close to normal distribution (see Figure 7.7 below). No reaction times of single trials were slower than 100 ms. Thus, no further steps of data preparation were required. For calculating Cronbach's alpha reliability coefficient, missing values of all remaining items and subjects were substituted by the mean value of each subject to avoid a small sample size due to listwise deletion.

Social Perception – SPf1

The number of missings due to the absence of a response was counted per item. Of the original 40 items, twelve items had more than 15 % missings. Surprisingly, five subjects had between 15 and 28 missing values in 28 items; they were excluded from further analysis. According to step 2 and 3, the reaction times of wrong responses and those preceded by a false alarm were set to missing. A false alarm could be interpreted as a possible distracter person in the video before the target person appeared. A wrong response represented a key stroke indicating the wrong target person in choice reaction time trials.

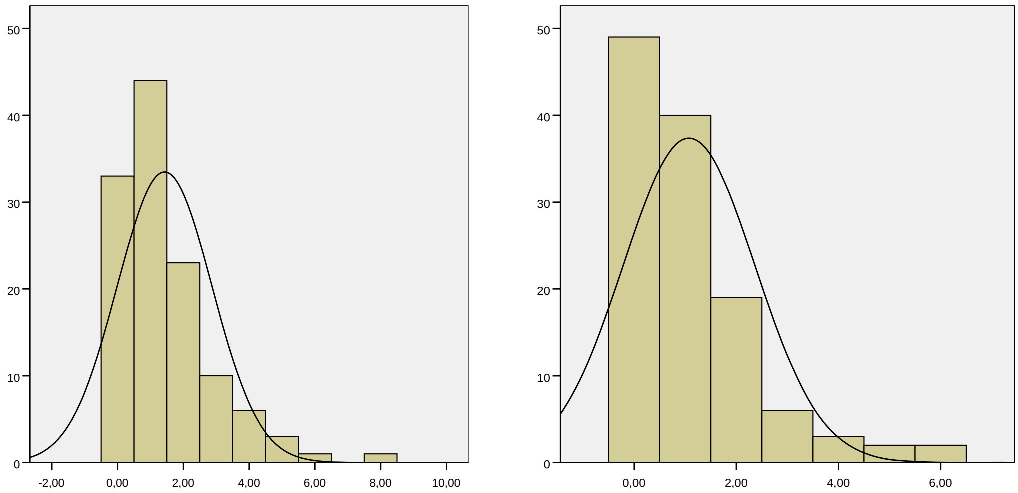


Figure 7.4

Histogram of Missing Values Due to False Alarms (Left Graph) and Wrong Responses (Right Graph) in SPf1 (Study 1)

The number of false alarms or wrong responses was low on average (see Figure 7.4; $m_{\text{false alarm}} = 1.43$, $sd_{\text{false alarm}} = 1.44$; $m_{\text{wrong}} = 1.07$, $sd_{\text{wrong}} = 1.29$, respectively). However, three more items were affected by a number of false alarms too large to include them in the analysis. No reaction times were slower than 100 ms. A screening of the final score distribution of the remaining items showed a close to normal distribution (see Figure 7.7 below). Two slight outliers were kept in the data set and inspected for influences on the bivariate distributions. Thus, no further steps of data preparation were necessary. For calculating Cronbach’s alpha reliability coefficient, missing values were substituted by the mean value of each subject to avoid a small sample size due to listwise deletion.

7.4.2 Psychometric Properties and Descriptives – Research Questions 1A

The present Chapter serves two purposes. On the one hand, basic descriptive statistics and reliabilities of the tests applied in the study will be presented. On the other hand, the Chapter intends to focus on test development. This mainly includes an investigation of item difficulties and item-total correlations (r_{it}) to find out about adequate item formats. The present chapter includes all analyses in the context of research question 1A.

7.4.2.1 Social Understanding Tasks

Analysis Prior to Selection of Item Format

The social understanding tasks were arranged within four different scenarios. The aggregation of items belonging to one scenario was not primarily of interest as the units of analysis. In contrast, scales were constructed according to the material-related task contents to represent the cells of the design. Therefore, all items of one content domain were combined to one score. Additionally, the final personality ratings were aggregated to one scale. The left side of Table 7.4 presents the results of a first reliability analysis showing the underlying number of items (item count), range in item-total correlations, and Cronbach's alpha coefficients based on all item formats. All parameters are presented for the entire scale including all formats in the first lines in boldface. Parameters are as well presented for the three types of item formats *within* the entire scale. The analysis showed a large range in item-total correlations for all item formats and no sufficiently reliable scales (α between .403 and .618).

On the right side of Table 7.4, the same parameters are presented after item selection. The scale was optimized based on the item-total correlations so that higher reliabilities were achieved. Exceptions were allowed only when one item was needed to maintain heterogeneity and a balanced taxonomy. Although some item-total correlations were still low, sufficiently reliable scales could be built for most of the scales based on all item formats (α between .635 and .772). No item format seemed to have failed completely so that the final scales comprised all item formats except for the pictorial task.

Table 7.4

Reliability Analysis Social Understanding Tasks, all Item Formats, Target Scoring (Study 1)

		All items			Items selected based on r_{it}		
Content domain	**Item format**	**Item count**	**r_{it} range**	**Cronbach's α**	**Item count**	**r_{it} range**	**Cronbach's α**
SUv	**Altogether**	**64**	**-.593 - .434**	**.594**	**33**	**.052 - .511**	**.742**
	Rating	52	-.360 - .434		27	.062 - .511	
	MC	7 (1)	-.593 - .157		3	.371 - .421	
	Open	4	-.033 - .236		3	.052 - .282	
SUa	**Altogether**	**71**	**-.292 - .379**	**.452**	**33**	**.049 - .502**	**.680**
	Rating	58	-.292 - .379		24	.049 - .502	
	MC	1	.150		1	.380	
	Open	12	-.218 - .244		8	.055 - .255	
SUp	**Altogether**	**26**	**-.112 - .337**	**.403**	**10**	**.063 - .509**	**.635**
	Rating	15	-.112 - .329		6	.063 - .509	
	MC	8	-.027 - .337		4	.286 - .391	
	Open	3	-.080 - .018		-	-	
SUf	**Altogether**	**78**	**-.159 - .376**	**.618**	**40**	**.097 - .422**	**.772**
	Rating	69	-.159 - .376		38	.097 - .406	
	MC	5	-.063 - .239		1	.223	
	Open	4	-.035 - .215		1	.413	
SUps	**Rating**	**36**	**-.113 - .599**	**.581**	**21**	**.123 - .422**	**.675**

Note. SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

Table 7.4 also shows the conceptual failure to provide a balanced number of items within the content-related scales representing all three item formats (e.g., one item based on multiple choice format and 58 based on rating format in the spoken language scale). The number of items per format was too unequal to allow a direct comparison of the psychometric qualities of the separate scales based on different formats each. Instead, to provide answers to question 1A2 (i.e., influence of item format on the psychometric properties), scales based on rating formats alone were composed. Ratings-based scales were chosen because this was the only type of item format which provided a reasonable number of items. These scales were then compared to the scales based on all item formats (right side of Table 7.4).

Analysis Based on Rating Format Scales – Target Scoring

Table 7.5 presents the results of the reliability analysis based on rating format alone. The scales on the left side of the table were based on target scoring. These scales were optimized according to the aforementioned principles, separately for the two scoring methods (see group consensus scoring on the right side of the Table). The upper line of each content-domain (in boldface) shows the parameters after item selection. The ratings-based scales (the left columns in table 7.5) and the scales containing all three item formats (the right columns in Table 7.4) achieved comparable levels of reliabilities although the number of items was smaller in the ratings-based scales (e.g., α for the spoken language tasks was .680 for all item formats and .678 for only rating format). For this reason and due to the limited possibility of constructing separate scales of other item formats, the subsequent analysis relied on scales based on rating format only. Table 7.5 also shows that the range in item-total correlations was large in the unselected scales (between -352 and .599; see lower line of each content domain). To build internally consistent scales, a large numbers of items had to be excluded. The final scales show sufficiently strong reliability coefficients (α between .653 and .764).

Table 7.5. also presents the range of item difficulties (i.e., mean difference from target answer in single items) prior to and after item selection. The range did not change at the low end of item difficulties (i.e., easy items were maintained in item selection). A relatively large proportion of items at the higher end of the difficulty scale were excluded. Thus, in all five scales, the range of item difficulties was reduced due to item selection (e.g., from -3.409 to -2.509 at the higher end of difficulty for the spoken language scale).

Table 7.5

Reliability Analysis Social Understanding Tasks, Rating Formats Only, Target and Group Consensus Scoring (Study 1)

	Target scoring				Group consensus scoring			
	Item count	**r_{it} range**	**Cronbach's α**	**Range of item difficulties***	**Item count**	**r_{it} range**	**Cronbach's α**	**Range of means**
SUv**	**27**	**.039 - .552**	**.745**	**[-3.115; -.32]**	**38**	**.090 - .515**	**.769**	**[.192; .971]**
	52	-.352 - .415	.564	[-3.573;-.041]	52	-.150 - .469	.725	[.192; .943]
SUa**	**22**	**.042 - .568**	**.678**	**[-2.509; -.235]**	**24**	**.059 - .542**	**.662**	**[.174; .443]**
	58	-.282 - .394	.416	[-3.409; -.235]	58	-.126 - ;361	.510	[.173; .743]
SUp**	**5**	**.087 - .694**	**.653**	**[-2.984; -.361]**	**9**	**.094 - .466**	**.549**	**[.217; .627]**
	15	-.107 - .365	.459	[-3.372;-.361]	15	-.042 - .427	.479	[.217; .631]
SUf**	**38**	**.083 - .404**	**.764**	**[-3.31; -.172]**	**35**	**.083 - .397**	**.691**	**[.190; .869**
	69	-.135 - .364	.620	[-3.713;-.124]	69	-.156 - .368	.596	[.184; .864]
SUps **	**21**	**.123 - .422**	**.675**	**[-2.262; -.654]**	**33**	**.085 - .372**	**.698**	**[.228; .477]**
	36	-.113 - .599	.581	[-2.694;-.533]	36	-.001 - .374	.693	[.228; .478]

Note. * possible range [-5.00; .00], a higher score indicating better performance
** *upper line* of each ability domain printed in boldface indicates scales with a reduced item number after item selection, *lower line* indicating scales prior to item selection
SU = social understanding, v = written language, a = spoken language, p = pictures, f = video, ps = personality ratings

Analysis Based on Rating Format Scales – Group Consensus Scoring

In the right columns of Table 7.5, the reliability analysis for group consensus scoring is shown. The concept of group consensus scoring has been discussed as problematic. The present analysis was not aimed at seriously considering group consensus scoring as an alternative scoring option. Rather, it was interesting to investigate the effect of this scoring method on the psychometric properties. Obviously, the range in item-total correlations of the unselected items (lower line of each content domain) showed better item properties. Thus, scales prior to item selection showed larger reliability coefficients for group consensus scoring. After item selection, however, the reliabilities could not attain a higher level compared to target scoring although more items were maintained in the scales. The range of means of the single items basically remained the same prior to and after item selection.

Distributions

An inspection of the univariate and bivariate distributions revealed two problems concerning outliers. One person showed an outlying value only in the pictorial scale. The outlier was nearly 7 *sd* below the group mean (see Table 7.6 for the parameters involving the outlier, in parenthesis of SUp parameters). Any mistakes in the raw data or any syntax error were checked and no mistake was found. The subject did not conspicuously differ in any other task and so no reason for this outlying performance could be detected. This value was substituted by the mean performance in the remaining social understanding scales. No change in bivariate correlations was detected afterwards. Another subject showed strongly outlying values on three of the social understanding scales which biased the bivariate correlations up to $r_{diff} = .17$ between the auditory and pictorial scale (see Table 7.7 above the diagonal). This person was omitted from the analysis. The final scales are presented in Table 7.6, showing the final N, means, standard deviations, range of scores, and distribution parameters, both for target and group consensus scoring. The final N arose from the exclusion of four subjects due to missing values and one subject due to the influence of outliers on the bivariate distributions and thus the correlational results. The means of the target scoring scales varied slightly between the content domains (*m* between -1.05 (SUp) and -1.612 (SUf)). t-test for paired samples showed significant mean differences between all paired comparisons (alpha level was adjusted according to the number of t-test) except for the differences between SUa with SUv and SUps. Thus, the pictorial scale was the one with the lowest mean difficulty whereas the video-based scale was most difficult to accomplish. This finding, however, could not be interpreted any further because no comparison values were available. Table 7.6 shows the influence of the outlier in the pictorial scale (in parentheses) which was substituted for the final scale.

Table 7.6

Descriptive Statistics of Social Understanding Tasks Target and Consensus Scoring (Study 1)

	Task	Item count	M	SD	Range	Skewness	Kurtosis
TS	SUv	27	-1.325	.406	[-2.58; -.66]	-.688	.055
	SUa	22	-1.254	.377	[-2.43; -.56]	-.413	-.024
	SUp*	5	-1.05 (-1.078)	.486 (.574)	[-2.80 (-4.40); -.20]	-.905 (-2.091)	.908 (8.755)
	SUf	38	-1.612	.364	[-2.59; -.90]	-.408	-.120
	SUps	21	-1.213	.397	[-2.57; -.49]	-1.024	1.342
GCS	SUv	38	.411	.053	[.27; .51]	-.311	-.634
	SUa	24	.315	.047	[.19; .40]	-.348	-.291
	SUp	9	.404	.090	[.08; .55]	-.911	.798
	SUf	35	.339	.043	[.20; .42]	-.848	.702
	SUps	33	.324	.045	[.17; .42]	-.553	.306

Note. N = 121, * in parentheses are parameters before deletion of an outlying data point (≈ -7 *sd*)
SU = social understanding, v = written language, a = spoken language, p = pictures, f = video, ps = personality ratings
TS = target scoring, GCS = group consensus scoring

The distribution parameters were basically normal showing the influence of the outlier prior to substitution (SUp in parentheses). Figure 7.5 illustrates the univariate distributions of the final social understanding scales (target scoring) of the different content domains. Comparable distribution forms emerged for the personality rating scales and for all scales based on group consensus scoring.

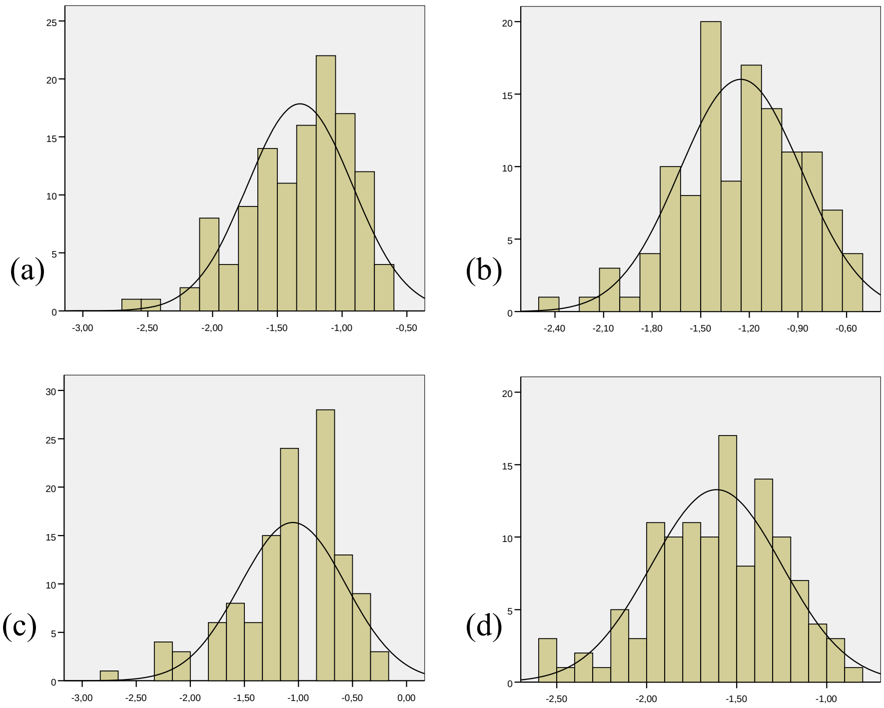


Figure 7.5

Histograms of Social Understanding Final Scales (Study 1)

Note. (a) SUv, (b) SUa, (c) SUp, (d) SUf

Correlations between Social Understanding Tasks – Target and Group Consensus Scoring

Finally, Table 7.7 shows the zero-order correlations between the social understanding scales within and between the scoring methods. Results yielded rather consistent correlations within one scoring method (r = .203 - .524 for target scoring) except for the scales based on personality ratings. They turned out as marginally correlated with the remaining scales (r = .037 - .116 for target scoring). Interestingly, the correlations within the target scoring scales based on all item formats were generally lower than those based only on rating formats (r = .120 - .469). These correlations are also presented in Table 7.7 (in parentheses in the upper half). This preliminarily pointed towards a method-related influence on the correlation size related to the different numbers of items of different formats. However, one additional finding put this finding into another perspective: The correlations between the personality rating scales (i.e., only ratings-based from the beginning) and the content-related scales based on all item formats were substantially higher for the written and the spoken language tasks while the remaining scales did not a show a change in correlation size. This could not have been expected because different item formats were underlying the scales. This result suggested that no method effect was responsible for the larger correlations within the content-related scales when only ratings-based format was applied. Thus, it strengthened the conclusion to rely on ratings-based items in the upcoming analysis and in test modifications. Still, it could not be clarified why the correlations with the personality rating scales were such small.

Table 7.7

Intercorrelations of Social Understanding Scales based on Target and Consensus Scoring (Study 1)

		Target Scoring					Group Consensus Scoring				
		SUv	**SUa**	**SUp**	**SUf**	**SUps**	**SUv**	**SUa**	**SUp**	**SUf**	**SUps**
Target Scoring	**SUv**		.563**	.351**	.509**	.102					
	SUa	.524** (.364**)		.373**	.501**	.053					
	SUp	.269** (.200*)	.203* (.120)		.324**	.080					
	SUf	.504** (.469**)	.509** (.331**)	.332** (.300**)		.033					
	SUps	.116 (.158)	.078 (.217*)	.127 (.098)	.037 (.004)						
Group Consensus Sc.	**SUv**	.748**	.454**	.125	.441**	.076					
	SUa	.422**	.864**	.108	.498**	.104	.456**				
	SUp	.127	.189*	.878**	.277**	.115	.200*	.307**			
	SUf	.506**	.614**	.171	.801**	.153	.386**	.554**	.396**		
	SUps	.113	.217*	.102	.100	.863**	-.041	.335**	.164	.247**	

Note. N = 121 below diagonal; N = 122 above diagonal
intercorrelations of scales based on all item formats in parentheses; correlations above the diagonal represent the parameters prior to the deletion of one subject because of multivariate outliers; shaded cells: intercorrelations between the scoring methods relying on the same content scales
SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

Target and group consensus scoring were highly intercorrelated between the corresponding content-related scales (r = .748 - .878, see lower part of Table 7.7). The remaining correlation coefficients between the methods were generally lower. These correlations, however, should not be interpreted any further as long as no empirical results were available that allow a conclusion about the nature of group consensus scoring (see Chapter 5.2.4 for more detailed considerations and Study 2 for further analyses).

7.4.2.2 Social Memory Tasks Based on Pictures and Videos

The social memory scales were inspected by the same kind of reliability analysis. It yielded rather low Cronbach's alpha coefficients (see Table 7.8; SMp1: $\alpha = .524$, SMf: $\alpha = .469$). Some items were not included in the final scales because of zero or negative item-total correlations (see second-left column of Table 7.8 for the number of items prior to and after item selection). The task SMp2 had already shown a large number of missing values. The remaining four items could not be aggregated to a compound score because of rather low item-total correlations. This task was omitted from any further analysis. Table 7.8 presents the descriptive statistics of the final pictorial and video-based tasks of social memory and perception including the reliability coefficients. The resulting reliability coefficients were not high. This was to be accounted for in the upcoming correlative analyses.

Table 7.8

Descriptive Statistics of Social Memory and Perception Tasks, Pictures and Videos (Study 1)

Task (N)	Item count	M	SD	Range	Skew-ness	Kurto-sis	r_{it} range	α
SMp1 (126)	13 (16)	.651	.176	[.23; 1.00]	-.159	-.766	.107 - .371 (-.052 - .339)	.524 (.484)
SMp2 (99)*	4	.000	.528	[-1.230; 1.110]	-.039	-.741	-.035 - .150	.136
SMf1+2 (115)*	9 (15)	.000	.437	[-1.13; .89]	-.447	-.167	.093 - 280 (-.151 - .236)	.469 (.335)
SPp1 (126)	74 (85)	1858.109	298.120	[1257.46; 2661.26]	.297	-.380	.121 - .734 (.038 – 509)	.897 (.873)
SPf1 (121)	23 (25)	4450.907	988.969	[2518.95; 8056.55]	.757	1.1666	.121 - .520 (-.121 - .520)	.697 (.683)

Note. * z-scores
in parentheses are parameter before item selection based on r_{it}-inspection
SM = social memory, SP = social perception, p = pictures, f = videos

Figure 7.6 displays the boxplots of the final score of the two memory tasks SMp1 and SMf1+2. The task SMp1 showed a slight ceiling effect. Item difficulties did not drop below the guessing rates but instead reached the high end of the scale (.432 - .904 for multiple choice items based on four choice alternatives). No distribution problems were encountered for the video-based task.

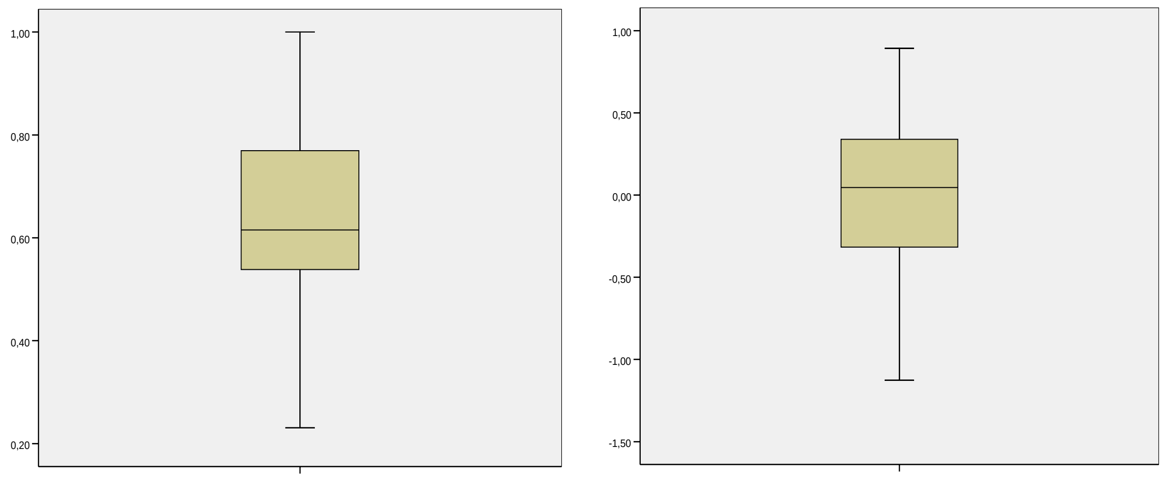


Figure 7.6

Boxplots of SMp1 (Left Graph) and SMf1+2 (Right Graph) (Study 1)

7.4.2.3 Social Perception Tasks Based on Pictures and Video

A reliability analysis suggested the exclusion of eleven and two items respectively for SMp1 and SMf1, due to zero or negative item-total correlations. The resulting reliabilities, however, were good. The descriptive statistics of the two social perception tasks based on pictures and videos are also presented in Table 7.8, including Cronbach's alpha prior to and after item selection. Long reaction times were observed in the two tasks. Consequently, the typical skewed distributions for reaction time scores was not found in the present study. Figure 7.7 presents the boxplots of the tasks showing a few outliers that were, so far, not trimmed in the data. This should only be done if the outliers turned out to bias bivariate distributions, which was not the case.

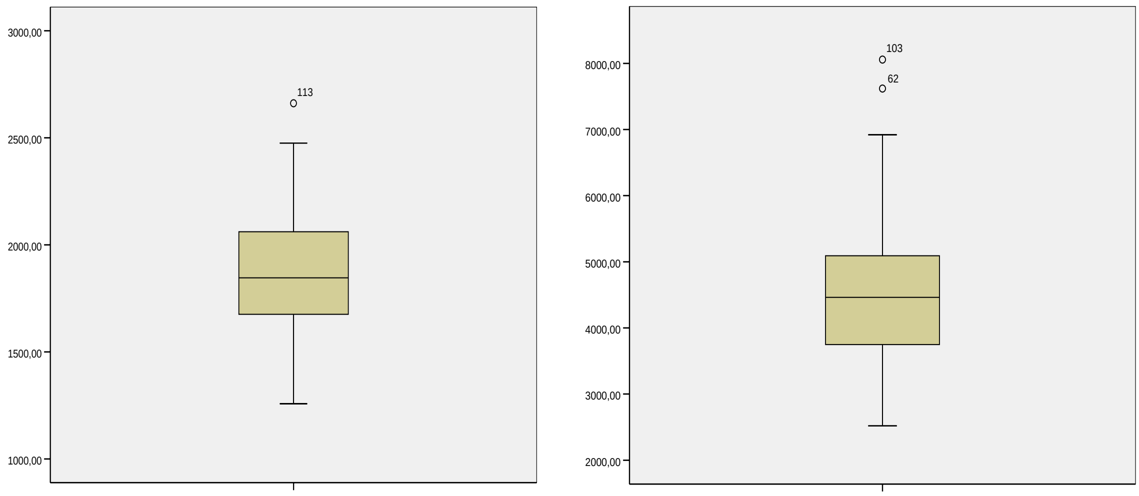


Figure 7.7

Boxplots of SPp1 (Left Graph) and SPf1 (Right Graph) (Study 1)

7.4.2.4 Further Measures

Table 7.9 presents the descriptive statistics and reliabilities of the social memory and perception tasks based on written and spoken language. The detailed analysis is presented in Seidel (2007). These parameters represent the final scores underlying further analysis in the present study. Any special problems encountered during scale construction were to be described in Seidel (2007).

Table 7.9

Descriptive Statistics of Social Intelligence Tasks Based on Written and Spoken Language (Study 1)

Task (N)	Item count	M	SD	Range	Skew-ness	Kurto-sis	r_{it} range	α
SMv1+2 (125)	26 (38)	.502	.13	[.06; .76]	-.646	.644	.16 - .46	.74
SMa1 (125)	61 (62)	.528	.092	[.21; .70]	-.921	1.609	-.05 - .45	.78
SPv1 (108)	32 (35)	5455.441	689.978	[3491.59; 7182.97]	-.282	.411	.11-62	.83
SPa1 (125)	109 (136)	792.507	68.45	[601.83; 956.69]	-.051	-.334	.10 - .46	.88

Note. in parentheses are item numbers before item selection based on r_{it}-inspection
SM = social memory, SP = social perception, v = written language, a = spoken language

Table 7.10 below presents the psychometric properties of the remaining tests and questionnaires applied in the present study and relevant for any research question. The BIS descriptives were based on the standardized aggregated cells for the respective operative and content-related ability domains. The reliability statistics for the operative and content-related ability domains were based on content- and operation-homogenous parcels, respectively. No special problems were encountered concerning the distributions or reliabilities. The reaction time data for the three baseline measures were treated according to the principle described in Chapter 7.4.1. Problems with missing values in the baseline measures did not occur. The distributions, however, were skewed, particularly for the SRT and the Readspeed task. These distributions were therefore modified as per steps 4 through 6 described in Chapter 7.4.1.

Table 7.10

Descriptive Statistics of Further Measures (Study 1)

Task (N)	Item count	M	SD	Range	Skew-ness	Kur-tosis	r_{it} range	α
BIS-R* (124)**	-	.000	.563	[-1.36; 1.32]	.015	-.434	.586 - .699	.813
BIS-M* (124)**	-	.000	.583	[-.156; 1.43]	.204	-.159	.503 - .526	.700
BIS-S* (124)**	-	.007	.561	[-1.33; 1.59]	.274	.026	.475 - .688	.723
BIS-V* (124)**	-	-.000	.586	[-.165; 1.31]	-.281	.381	.457 - .706	.726
BIS-F* (124)**	-	.000	.498	[-1.17; 1.20]	-.101	-.337	.414 - .545	.646
BIS-N* (124)**	-	.007	.593	[-1.16; 1.80]	.514	.093	.651 - .709	.815
SRT (125)	50	242.140	21.453	[203.60; 300.36]	.504	.082	.415 - .720	.961
MT (126)	75	686.280	69.354	[538.33; 885.30]	.339	-.052	.344 - .725	.976
Readspeed* (122)	198	.000	1.000	[-2.253; 3.013]	.226	.249	.590 - .849	.996
NEO-N (126)	12	1.682	.555	[.67; 3.17]	.351	-.470	.241 - .613	.808
NEO-E (126)	12	2.441	.480	[1.33; 3.33]	-.246	-.584	.149 - .697	.755
NEO-O (126)	12	2.525	.498	[1.25; 3.92]	-.103	-.102	.082 - .552	.716
NEO-A (126)	12	2.540	.451	[1.42; 3.42]	-.299	-.265	.070 - .582	.722
NEO-C (126)	12	2.631	.588	[1.00; 3.92]	-.302	-.338	.305 - .667	.863
SB Questionnaire (123)	40	2.701	.330	[1.98; 3.78]	.152	.390	.135 - .536	.825

Note. * z-scores
** reliability analysis based on content- and operation-homogenous parcels, four parcels for BIS-R, three parcels for BIS-S, -M, -V, -F, -N
R = Reasoning, M = Memory, S = Speed, V = verbal abilities, F = figural-spatial abilities, N = numerical abilities, NEO-N = Neuroticism, NEO-E = Extraversion, NEO-O = Openness, NEO-A = Agreeableness, NEO-C = Conscientiousness, SB = social behavior, α = Cronbach's alpha

7.4.3 Construct Validity

The research questions underlying the following analysis were twofold. First and foremost, the structure of social intelligence assessed by the newly constructed tasks was examined. Secondly, the relationship between social intelligence and academic intelligence and personality traits was investigated in order to prove divergent construct validity.

7.4.3.1 Structure of Social Intelligence – Research Question 2A

Correlational Results Based on Original Scales

The research question underlying the present analysis concerned the internal structure of social intelligence as assessed by the newly developed tasks. Social intelligence and the more specific operative ability domains were to prove consistently positive within-domain correlations and should load on the same factors. Prior to any multivariate analysis, Table 7.11 presents the correlations between the newly developed tasks of social intelligence. The social understanding and social memory operative domains showed coherent within-domain correlations. The social understanding tasks correlated with $r = .204$ - $.524$ with each other, the lowest correlation was between the spoken language and the pictorial task. The social memory tasks showed correlations between $r = .096$ and $r = .526$. Again, the lowest correlations were associated with the pictorial task. In turn, the highest intercorrelations were found for language-based tasks. This finding, however, could be associated with the rather low reliability coefficient of the pictorial and the rather high coefficients of the language-based tasks.

The correlations within the social perception domain, however, did not support a coherent ability domain. Only the pictorial and video-based tasks showed substantial convergent overlap. The pattern of correlations across the ability domains was unsystematic. Particularly the auditory social memory task was substantially correlated with the social understanding tasks. No further tasks showed systematic overlap with tasks related to other ability domains. Tasks based on common contents showed only partially coherent within-domain correlations. The language-based tasks especially, seemed to share some common variance. However, any further evidence about the internal structure of social intelligence and support for the different ability domains should be derived from confirmatory factor analysis.

Table 7.11

Correlations Between Social Intelligence Tasks (Study 1)

	SUv	SUa	SUp	SUf	SMv1	SMa1	SMp1	SMf1	SPv1	SPa1	SPp1	SPf1
SUa	.524**											
	(121)											
SUp	.269**	.204*										
	(120)	(120)										
SUf	.504**	.509**	.333**									
	(121)	(121)	(120)									
SMv	.248**	.132	.160	.220*								
	(120)	(120)	(119)	(120)								
SMa1	.296**	.205*	.247**	.191*	.526**							
	(121)	(121)	(120)	(121)	(124)							
SMp1	.066	.107	-.009	-.060	.096	.179*						
	(121)	(121)	(120)	(121)	(125)	(125)						
SMf	-.139	-.002	.053	-.060	.269**	.342**	.240**					
	(110)	(110)	(109)	(110)	(114)	(114)	(115)					
SPv1	.014	-.064	-.103	.013	-.248*	-.081	-.118	-.065		.037	.185	-.053
	(104)	(104)	(103)	(104)	(107)	(108)	(108)	(98)		(103)	(104)	(101)
SPa1	.006	-.052	.006	.093	.021	.063	.043	-.041	.004		.042	.034
	(120)	(120)	(119)	(120)	(124)	(124)	(125)	(114)	(107)		(120)	(116)
SPp1	-.104	-.117	.136	-.040	.023	.078	-.046	-.070	.183	.100		.409**
	(121)	(121)	(120)	(121)	(125)	(125)	(126)	(115)	(108)	(125)		(116)
SPf1	-.025	-.048	.055	-.064	-.035	.109	.023	-.024	-.076	.126	.471**	
	(117)	(117)	(116)	(117)	(120)	(121)	(121)	(111)	(105)	(121)	(121)	

Note. pairwise N in parentheses; * p < .05; ** p < .01
correlations between social perception tasks corrected for speed baseline above diagonal
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

Correcting for Speed Baseline Variance in Social Perception?

Thus far, the social perception tasks represented the original scales. Initially, it was intended to control for the relevant mental speed baselines. Therefore, three baseline speed measures had been applied in the present study, assessing reading speed as baseline for the verbal social perception task (SPv1), simple reaction time task (SRT) as baseline for tasks relying on keystrokes (SPv1, SPa1, and SPf1), and mouse speed task (MT) as baseline for the pictorial social perception task (SPp1). Additionally, it was thought that SRT might also influence the task SPp1 (based on the use of the mouse) because it included the most basic mental speed requirements not primarily included in the MT which had more coordinative requirements.

The social perception tasks correlated with the baseline measures to varying degrees. As expected, SPv1 showed the highest correlations with the readspeed baseline (r = .212, $p < .05$, N=104) and zero correlations with both other baseline measures. SPa1 was substantially related to SRT (r = .371, $p < .01$, N=124). SPp1 correlated most highly with the mouse speed task (r = .397 $p < .01$, N=126), SPf1 correlated with the SRT (r = .187, $p < .05$, N=120). However, against expectations, the SPf1 task showed a larger correlation with the mouse speed baseline (r = .231, $p < .05$, N=121) than that with the intended SRT baseline measure. The baseline measures themselves correlated only marginally positively with one another with only one significant correlation between the SRT and Readspeed (r = .218, $p < .05$, N=121).

Because of the equivocal correlation pattern between the baseline measures and the social perception tasks, all common speed baseline variance was partialled out of the social perception tasks in a multiple regression analysis and the residuals were saved as baseline corrected scales. However, controlling for the baseline variance did not change the correlation pattern within the social perception domain (see Table 7.11 above diagonal). The picture- and video-based tasks still correlated substantially (r = .409) and the relationship between SPv1 and SPp1 remained nearly exactly the same (r = .185). Most of the other correlations lost in size so that a coherent social perception domain was to be questioned. To anticipate the further analysis, structural equation modeling could not support a measurement model of social perception so that this domain was not included in further analysis about the internal structure of social intelligence. In any later analysis, only the baseline corrected scales were applied.

Confirmatory Factor Analysis

The SIM relied on the performance model of social intelligence (Weis & Süß, 2005). Consequently, the hypotheses postulated a structural model of social intelligence with three correlated ability factors as well as a hierarchical model of social intelligence, reflected in a Schmid-Leiman solution. Confirmatory factor analysis was believed to provide the required empirical support for the postulated structure of social intelligence. To test the hypotheses, several models were postulated. Table 7.12 presents the rational underlying the tested models and the summary of fit statistics. The models underlying the analysis postulated a general social intelligence factor with loadings of all social understanding and memory variables (Model A), a structural model with two correlated operative ability factors, social understanding and memory (Model B), a structural model with two uncorrelated operative ability factors (Model C), and a hierarchical social intelligence model with two uncorrelated operative ability factors and one general social intelligence factor with loadings of all variables (Schmid-Leiman solution) (Model D)

Before turning to the results, two additional remarks are necessary. *(a)* Prior to the factor analysis, measurement models for the single operative ability domains were established. These supported social understanding and social memory as coherent ability factors with positive and meaningful loadings of all indicators. A measurement model for the social perception factor could not be established involving all indicators. *(b)* Due to the exclusion of several subjects in different tasks, the listwise N was only 109. This number was clearly at the lower end of the possible sample size to analyze the postulated models. Therefore, all conclusions should be derived with care and needed replication in Study 2. No special problems, however, were encountered during the analysis except for model D. Due to a condition code because of an error term at lower bound, an equality constraint was introduced.

Table 7.12

Fit Statistics for Confirmatory Factor Analyses of the Structure of Social Intelligence (Study 1)

Model	χ^2	DF	p (χ^2)	CFI	RMSEA	SRMR	CI RMSEA*
A: General factor model	53.904	20	<.001	.750	.125	.103	[.085; .165]
B: 2-Factor structural model (SU and SM correlated)	26.391	19	.119	.946	.060	.071	[.000; .110]
C: 2-Factor structural model (SU and SM uncorrelated)	33.464	20	.030	.901	.079	.103	[.025; .124]
D: Hierarchical model (Schmid-Leiman)**	9.436	13	.739	1.000	.000	.037	[.000; .070]

Note. N = 109; * CI = 90%; ** two error terms constrained to be equal
SU = social understanding, SM = social memory, p = pictures, f = videos

Table 7.12 presents the fit statistics of the confirmatory factor analysis for all postulated models. The general factor model (Model A) did not show a good data fit (CFI = .750; χ^2 = 53.904, p < .001). The loadings of the pictorial and video-based social memory tasks were close to zero while the remaining tasks all loaded positively on this factor. The structural model with two correlated operative ability domains (Model B) showed reasonable fit statistics (CFI = .946; χ^2 = 26.391, p=.119). Figure 7.8 displays the standardized solution showing factor loadings and the factor intercorrelation. Error terms are displayed besides the manifest variables. Except for the pictorial social memory task, all tasks loaded significantly on the postulated ability factors (between .34 and .78). Factors were substantially correlated (r = .35). Model C tested the same structural model with uncorrelated ability factors. Data fit was weaker than that for model B (CFI = .901; $\chi^2 = 33.464$, $p < .05$). Models B and C were nested. A χ^2-differences test showed a significant increase of model fit for Model B with correlated ability factors (χ^2-difference = 7.037, df = 1, $p < .01$).

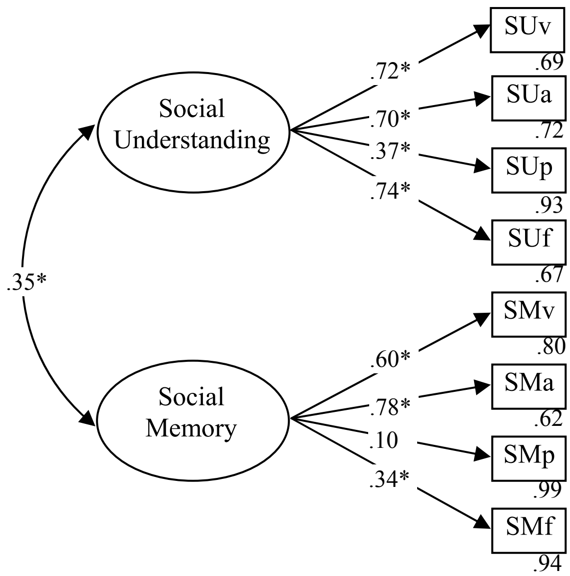


Figure 7.8

Standardized Solution of the Structural Model of Social Intelligence (Model B) with Two Operative Ability Domains (Study 1)

Note. CFI = .946; $\chi^2 = 26.391$ (p=.119); * $p < .05$; error terms are enclosed with the manifest variables
SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

Model B and the χ^2-differences test comparing Model B and C supported the convergence between the two operative ability factors and thus, the existence of a higher-order general social intelligence factor. This should be tested in the next model. A hierarchical Schmid-Leiman solution was postulated in Model D. As was mentioned before, some problems were encountered during the analysis so that two error terms were constrained equal (see Figure 7.9). Thus, the results of the analysis should be interpreted with care and the model needed replication in the next study. Fit statistics, however, turned out to be very good

(CFI = 1.000; χ^2 = 9.436, p=.739). The standardized solution of the model is shown in Figure 7.9. The loadings on the two uncorrelated operative ability factors were heterogeneous but were all in one direction (between -.27 and -.74 on the social understanding factor and between .22 and .89 on the social memory factor). The loadings of the social understanding tasks on the respective factor were all negative. Since the factor was not correlated with any other factor, the negative loadings did not present a problem. Interestingly, the pictorial task of social memory showed a higher loading on the social memory factor than in the structural model which was also significant (.22 compared to .10). Except for the pictorial and video-based tasks of social memory, all tasks loaded positively on the general social intelligence factor. Only the loadings of SUv, SMv, and SMa reached significance.

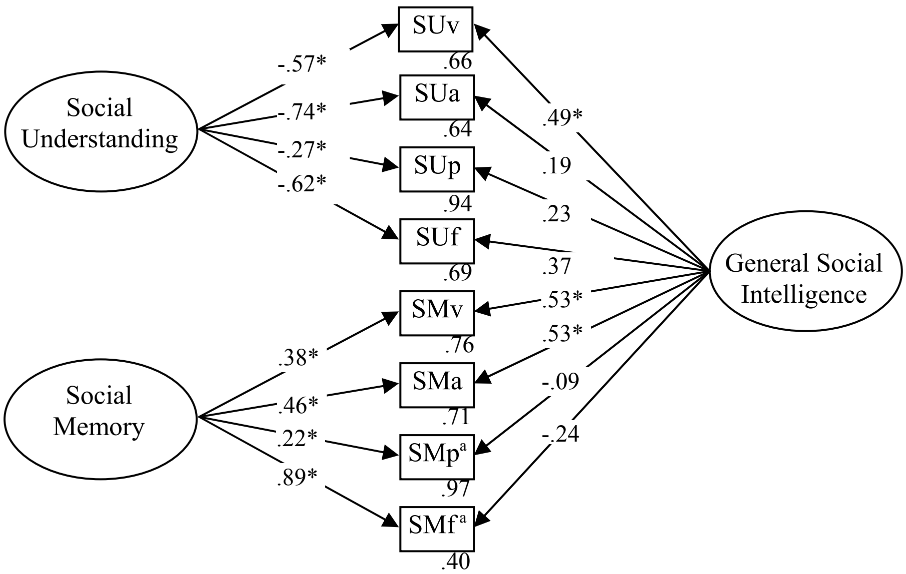


Figure 7.9

Standardized Solution of the Hierarchical Model (Schmid-Leiman; Model D) of Social Intelligence (Study 1)

Note. CFI = 1.000; χ^2 = 9.436, p=.739; * p < .05; error terms are enclosed with the manifest variables
[a] error terms constrained to be equal
SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

Originally, it was intended to explore the content-related domains as possible meaningful ability factors. The design of the SIM applied in the present study, however, was restricted and not all cells in the design were operationalized by two tasks. Moreover, the social perception tasks had not shown any convergent validity evidence so that these were not applied. Thus, it was not possible to investigate the structure of the content domain relying on factors related to only one content domain because these would have relied on two tasks only. Therefore, this analysis should be focused on in the next study that included a more elaborate design of the SIM. Furthermore, the analysis of the faceted structure of social intelligence was also postponed to Study 2.

In summary, the results supported the postulated structure and pointed towards a hierarchical model of social intelligence. At the same time, confirmatory factor analysis reflected the aforementioned problems related to the lack of convergence of the social perception domain and to the lack of reliability of pictorial and video-based tasks of social memory.

7.4.3.2 Divergent Construct Validity – Research Question 2C

Relationship to Academic Intelligence – Research Question 2C1

The study aimed at proving that social intelligence as assessed by the newly developed tasks could be separated from academic intelligence. Academic intelligence was measured by the BIS-Test (Jäger et al., 1997) which allowed an investigation of the construct relationship on different hierarchical levels (i.e., a general factor level, the level of broad ability domains related to operations and contents, and a more specific level of the cells resulting from the cross-classification of the operative and content-related ability domains). The single tasks of social intelligence most closely reflected the hierarchy level of the single cells of the BIS-Test. Therefore, as a first step, the intercorrelations of the tasks to the BIS cells were inspected (Table 7.13).

The four social understanding tasks were only marginally correlated with all BIS cells (r between -.172 and .215 with most correlations around zero). The social memory tasks correlated substantially with cells of all three BIS operative ability domains except for the video-based task. Particularly, the correlations of the tasks based on written and spoken language with BIS-Reasoning and BIS-Speed were equally large or even larger than the correlations with the BIS-Memory ability domain (Table 7.13). Thus, correlations generally suggested large overlap of the social memory domain with BIS operative ability domains. However, no clear pattern of overlap was discovered so that further analysis should rely on the investigation of the factor correlations.

Table 7.13

Correlations of Social Intelligence Tasks with BIS Cells (Study 1)

	BIS cells								
	RV	**RF**	**RN**	**MV**	**MF**	**MN**	**SV**	**SF**	**SN**
SUv (119)	.203*	.026	.038	.068	.091	.078	.101	-.037	.007
SUa (119)	.082	-.056	.051	.153	.081	-.086	.215*	-.110	-.034
SUp (119)	.040	-.107	.009	.058	-.027	.002	.077	.007	.000
SUf (119)	-.012	-.053	-.086	.106	.108	-.040	.062	-.085	-.172
SMv (123)	.456**	.178*	.240**	.323**	.067	.102	.489**	-.192*	.184*
SMa1 (123)	.446**	.042	.202*	.282**	.193*	.205*	.503**	-.150	.171
SMp1 (124)	.252**	.010	.153	.191*	.223*	.246**	.141	.016	.184*
SMf (113)	.101	-.048	-.016	.079	.045	.124	.276**	-.143	.085
SPv1 (103)	-,244*	-,196*	-,125	-,257**	-,021	-,115	-,311**	,017	-,227*
SPa1 (119)	,080	,084	,076	-,212*	-,077	,013	-,064	-,129	-,062
SPp1 (120)	-,125	,008	,005	-,044	,009	,051	-,088	-,164	-,042
SPf1 (115)	-,005	-,040	,055	,019	-,113	-,166	-,076	-,149	-,003

Note. pairwise N in parentheses; * p < .05; ** p < .01
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings, R = BIS-Reasoning, M = BIS-Memory, S = BIS-Speed, V = verbal, F = figural-spatial, N = numerical

The social perception scales in the present analysis represented the baseline corrected scales. The speed baseline measures themselves showed partly substantial correlations with the BIS cells (i.e., SRT: *r* between -.256 and .063; MT: *r* between -.382 and -.100; Readspeed: *r* between -.313 and .164). Therefore, the baseline corrected scales of social perception were applied in order to investigate the pure overlap with the BIS cells without common speed baseline variance. Results showed still meaningful correlations of the written language social perception task (SPv1) with all three BIS operative domains and particularly, with cells based on verbal material. In contrast, the correlations of the remaining social perception residual variables with the BIS domains were generally around zero. Thus, no further conclusions about the measurement constructs of the so could be derived from this result.

Correlational analysis provided first evidence about the overlap of social and academic intelligence and suggested an independent social understanding domain. Results related to the social memory domain did not allow such a clear interpretation. Moreover, the BIS cells, were not independent from one another (*r* ranging from .011 for cell MF with SF to .694 for cell RN with SN) so that further evidence about construct overlap was to be derive from confirmatory factor analysis.

Prior to investigating the relationship between social and academic intelligence based on the latent factor intercorrelations in the confirmatory factor analysis, the fit of the BIS structure as the model of reference was investigated by structural equation modeling. Therefore, two models were postulated. The models were based on content- and operation-homogeneous parcels, respective for the operative and content abilities. The parcels for the operative ability domains consisted of equal variance components from every content domain (i.e., content-homogeneous). In turn, the operation-homogeneous parcels for the content abilities consisted of equal variance components for each operative ability. Model E included three correlated operative ability factors (BIS-Reasoning, -Memory, -Speed, see Table 7.14) based on content-homogenous parcels. The second model contained three correlated content ability factors (BIS-Verbal, -Figural-Spatial, -Numerical, see Model F in Table 7.14) based on operation-homogenous parcels. Data fit from confirmatory factor analysis supported both models (see Table 7.14 Models E and F). Every parcel loaded positively and significantly on the respective factors, and factor intercorrelations were consistently high. The BIS structure had already been replicated several times and therefore an illustration of the BIS structure as supported by confirmatory factor analysis is not presented at this point.

Thus, the models of reference for investigating the construct overlap of social and academic intelligence were *(a)* the two-factor structural model of social intelligence with correlated ability factors and *(b)* the a model of the corresponding operative ability factors BIS-Reasoning and BIS-Memory. It was refrained from investigating the fit of models relying on general factor or hierarchical models. The general factor model of social intelligence had shown bad data fit. Moreover, relying on a hierarchical model on both sides to investigate construct overlap would have implied too many parameters for the comparably small sample size. Table 7.14 presents the fit statistics.

Table 7.14

Fit Statistics for Confirmatory Factor Analyses of the Divergent Construct Validity of Social with Academic Intelligence (Study 1)

Model	χ^2	DF	p (χ^2)	CFI	RMSEA	SRMR	CI RMSEA*
E: BIS-operative factors[a]	42.250	32	.106	.971	.052	.050	[.000; .091]
F: BIS-content factors[a]	26.065	24	.349	.994	.027	.047	[.000; .081]
G: SI-BIS 4 correlated ability factors[b]	100.049	84	.112	.954	.043	.075	[.000; .071]
H: SI-BIS, SU uncorrelated with BIS[b]	101.229	86	.125	.956	.041	.077	[.000; .070]
I: 2-Factor structural model (residuals of SU and SM, correlated)[c]	21.702	13	.06	.920	.079	.076	[.000; 136]

Note. * CI = 90%; [a] N = 119; [b] N = 106; [c] N = 107
SU = social understanding, SM = social memory, p = pictures, f = videos

Initially, all possible factor intercorrelations were postulated in Model G (i.e., between social understanding and memory, between BIS-Reasoning and –Memory, and between all BIS and social intelligence factors). The model showed reasonable data fit (CFI = .954; χ^2 = 100.049, p=.112). The factor loadings were consistently high. Meaningful and significant factor intercorrelations between social memory and both BIS factors, as well as between social memory and social understanding were observed. Social understanding showed a true zero correlation with BIS-Reasoning (r = .00) and only a marginal relationship with BIS-Memory (r = .15). Therefore, the second model dropped the correlations between social understanding and both BIS-factors (Model H). The model fit was not meaningfully better (CFI = .956; χ^2 = 101.229, p=.125). Models G and H were nested, the χ^2-differences test was not significant (χ^2-difference = 1.18, df = 2, n.s.). In accordance with the principle of parsimony, the more restricted model (Model H) was accepted proving the independence of social understanding from the BIS-Memory and the BIS-Reasoning domain. Figure 7.10 displays the standardized solution of Model H showing the remaining factor intercorrelations and the factor loadings. The loadings on the social intelligence ability factors equaled the loadings from the two-factor structural model of social intelligence (see Figure 7.8). Moreover, the correlation between the two social intelligence factors stayed the same. The social memory factor, however, correlated with the BIS ability factors to a larger extent than with social understanding. Another model, however, that constrained the factor intercorrelations between social memory and BIS factors to 1, did not converge in confirmatory factor analysis. Therefore, the independency of the factors could not be tested at this stage.

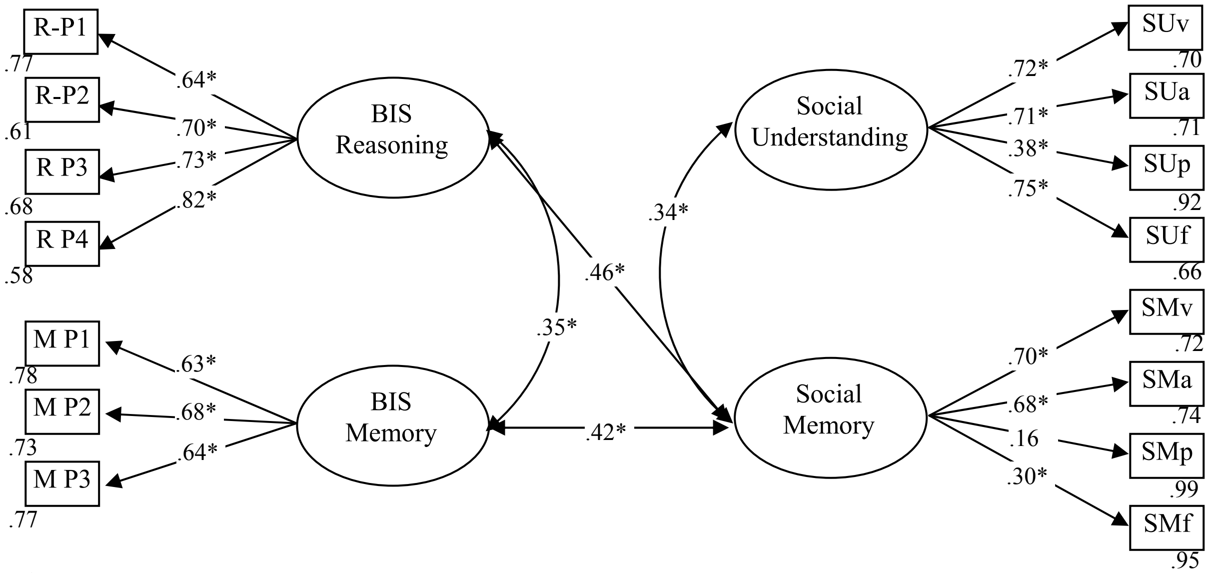


Figure 7.10

Standardized Solution of Construct Overlap of Social and Academic Intelligence (Model H) (Study 1)

Note. CFI = .956; χ^2 = 101.229, p=.125; *p < .05; error terms are enclosed with the manifest variables
R = BIS-Reasoning, M = BIS-Memory, P1-3 = Parcel 1-3, SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

The preceding analysis could not convincingly prove whether social memory was independent from academic intelligence, particularly from BIS-Memory. Therefore, as a last step to prove divergent construct validity with academic intelligence, the previously identified structure of social intelligence should show independence from the BIS structure. Therefore, the structural model of social intelligence was supported by previous analysis was tested once again by confirmatory factor analysis, this time relying on the residuals of the single tasks after partialling out the complete BIS variance (i.e., BIS variance of all operative and content scales was entered in multiple regression analysis predicting the single social intelligence tasks; the residuals of the social intelligence tasks were saved). Figure 7.11 presents the standardized solution of the model, fit statistics are presented in the last row of Table 7.14. The model fit dropped slightly compared to the model based on the original data (CFI = .920; χ^2 = 21.702, p=.06), however, the factor loadings and factor intercorrelations were obtained in comparison to the initial model based on the original variables. Only the pictorial social memory task did not fit into the model with a zero loading on the social memory factor. Therefore, the loading was omitted from the model. This last model provided first evidence for the structural independency of both social intelligence ability factors from the structure of the BIS despite the, in part, substantial between-domain correlations, particularly for social memory.

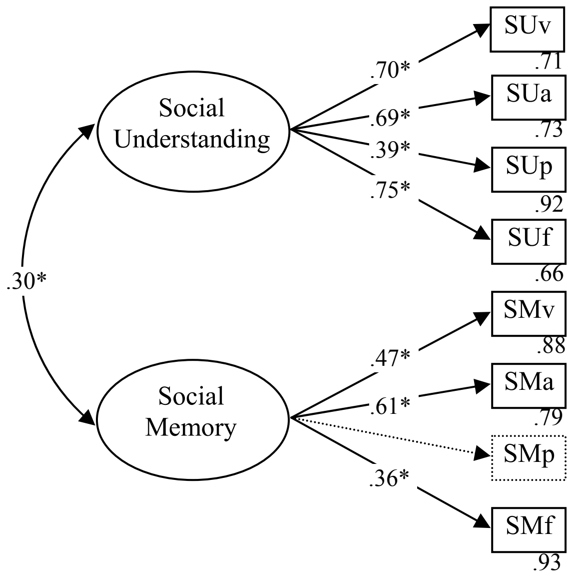


Figure 7.11

Standardized Solution of Two-Factor Structural Model of Social Intelligence When BIS Variance was Controlled (Model I) (Study 1)

Note. CFI = .920; χ^2 = 21.702, p=.06; * p < .05; error terms are enclosed with the manifest variables
SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

Relationship to Personality Traits – Research Question 2C1

As hypothesized, the tasks of social intelligence did not show substantial correlations with the Big Five personality traits assessed by the NEO-FFI (Borkenau & Ostendorf, 1993). Table 7.15 presents the correlations. Only the Openness factor correlated significantly with social memory tasks. The size of the correlations, however, did not raise doubts about the divergent construct validity of social intelligence from personality traits.

Table 7.15

Correlations of Social Intelligence Tasks with Big Five Personality Traits (Study 1)

	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C
SUv (121)	.050	-.167	.004	.062	-.058
SUa (121)	.149	-.035	.031	.085	.024
SUp (121)	-.027	-.059	-.067	-.018	.013
SUf (121)	.060	.047	.022	.075	.032
SMv (125)	.007	-.047	.177*	-.031	-.047
SMa1 (125)	-.009	-.069	.180*	.099	.039
SMp1 (126)	.045	.150	.178*	.075	.036
SMf (115)	-.089	.112	.131	.088	.101
SPv1 (108)	,016	-,079	-,185	,030	,078
SPa1 (125)	,046	,011	,130	,132	-,014
SPp1 (126)	-,089	-,099	,001	,018	,023
SPf1 (121)	-,013	-,153	-,102	,013	-,085

Note. pairwise N in the left column; * p < .05
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos, NEO-N = Neuroticism, NEO-E = Extraversion, NEO-O = Openness, NEO-A = Agreeableness,
NEO-C = Conscientiousness

7.4.4 Further Exploratory Questions – Research Questions 3A / 3B

Relationship to Self-Report Data – Research Question 3A

Correlations between the newly developed tasks and self-reported socially intelligent behavior (Amelang et al., 1989) were marginal with only one significant negative correlation with the spoken language scale of social understanding ($r_{SUa;self\text{-}report} = -.235$; $p < .05$; N = 118). This conformed, however, to frequently reported empirical results in existing literature (Brown & Anthony, 1990; Riggio et al., 1991) and to the expectations of the researcher.

Gender Differences – Research Question 3B

Gender differences in favor of women have been found to be interpreted as a positive validity result for tests intended to assess social and emotional abilities (Hall, 2001; Schutte et al., 1998). It would need additional theoretical underpinnings to support this idea which was not intended to be accomplished in the present work. Therefore, a specific direction of gender differences was not hypothesized and the following analysis was only exploratory in nature. To avoid inflation of the alpha-probability, the alpha level was adjusted dividing the alpha-probability (.05) by the number of exploratory tests (12 tests related to 12 social intelligence tasks), which resulted in a new alpha-probability of .004.

A t-test for independent samples showed gender differences of which nearly all were in favor of women (except for two tasks). Only two video-based tasks, however, showed a significant effect (SUf: t =-3.617, df = 119, p < .001, Cohen's d = .629; SMf: t = -3.109, df = 113, p=.002, Cohen's d = .563). Two more tasks showed meaningful effect sizes also in favor of female subjects without reaching statistical significance (SUa: Cohen's d = .464; SMa: Cohen's d = .457).

Gender Differences in Targets, Subjects, and Their Interaction – Research Question 3B1

Bronfenbrenner et al. (1958) found gender effects on judgmental accuracy, depending on the target's and the subject's gender, and on their interaction. To bring his results back to memory (see Chapter 4.3.3.5), he reported a positive correlation between performance of male subjects in judging male and males judging female targets; and a negative correlation between performance of female subjects in judging male and females judging female targets. He attributed this to a similarity effect of females targets with an better performance when judging targets of the same sex. His study, however, did not include the same targets for every judge so that any inference or comparison to different targets was not possible. This question should be addressed in the upcoming analysis.

A general social understanding scale for every target was built by aggregating the items of the final scales across the different content domains excluding the personality ratings since they were not correlated with the remaining scales. Both scales for female targets (SU_KL and SU_RF) and for male targets (SU_CP and SU_MM) were combined to form one score each. Both scores were highly correlated (r = .539, p < .01, N=121). As a first step, the same correlative analysis, as in the Bronfenbrenner study, was conducted . Results showed substantial positive correlations for both gender groups of subjects with r = .478 of (males judging males) with (males judging females) and r = .541 of (females judging males) with (females judging females). This did not conform with Bronfenbrenner's findings. Similarity between subjects' and targets' gender did not seem to have an impact on the performance.

Secondly, a two-factor analysis of variance (repeated measurements) was conducted postulating one between-subjects factor (i.e., the subject gender) and one repeated measurement factor (target gender). Results are presented in Table 7.16. Both main effects were highly significant showing women outperforming men in judging both male and female targets and female targets being easier to judge than male targets. No significant interaction effect was discovered. Figure 7.12 illustrates the results by the help of a graph. Again, the present results did not suggest an influence of the similarity in gender.

Table 7.16

Results of Two-Factor Analysis of Variance Examining the Effect of Target and Subject Gender on Performance (Study 1)

Source of variance	Sums of square	DF	Mean squares	F	Probability	Eta²
Repeated measures (target gender)	2.269	1	2.269	40.409**	<.001	.072
Between-subjects (subject gender)	1.899	1	1.899	11.009**	<.001	.060
Interaction effect	.071	1	.071	1.264	.263	

Note. N = 122

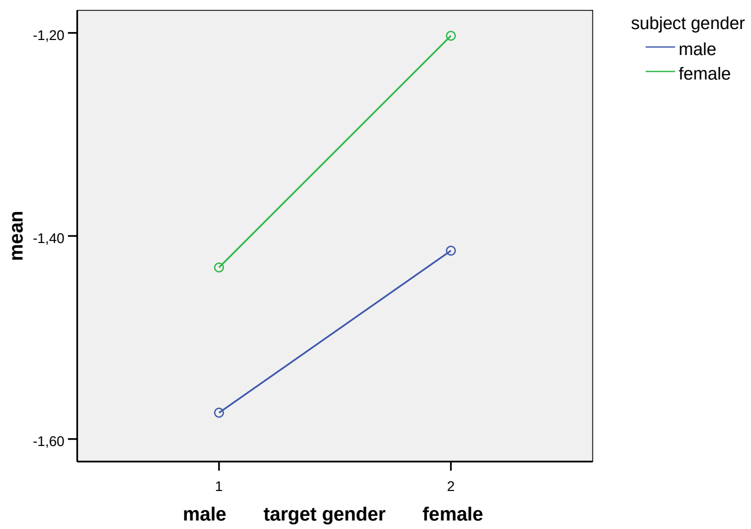


Figure 7.12

Results of the ANOVA Showing Gender Effects on Performance in Social Understanding Scales (Study 1)

These outcomes, however, were only based on two male and two female targets and should be replicated in Study 2. Thus, no generalization was possible. More importantly, different targets were judged on different items so that this result could as well be an effect of items of different difficulties underlying the scales. The controversial results of Bronfenbrenner and of the present study demonstrate the importance of conducting further studies containing different and, above all, numerous targets so that any effect of single target persons can be reduced. If the results could be replicated, this could mean different things. For example, women could be generally easier to judge (i.e., a higher sending accuracy) or their target answers could correspond better with the cues contained in the stimulus material. These questions, however, cannot be answered at this point.

7.5 Summary and Discussion Study 1

In general, this first empirical attempt to investigate the psychometric properties and validity of the SIM showed promising results. The reliabilities of most of the tasks were sufficiently high after item selection. The number of items that showed low item-total correlations was generally low, however, this varied between the tasks and was largest for social understanding. The results of construct validation supported the postulated structure of social intelligence with two correlated ability factors - social understanding and memory. All tasks loaded on their respective factors with, however, low loadings of the pictorial and video-based tasks of social memory which could be attributed to the rather low reliability coefficients for those tasks. The structural model was replicated relying on the residuals of the social intelligence tasks when BIS variance was partialled out showing structural independency from the BIS structure. Although with slight psychometric problems, a hierarchical model of social intelligence was supported by the data. This model clearly needed replication in the second study. Correlational results and confirmatory factor analysis clearly demonstrated an independent social understanding factor. Social memory was substantially correlated with BIS-Memory. However, a model postulating perfectly correlated memory factors for both constructs (i.e., fixing the factor intercorrelation to $r = 1$) did not converge and should be tested in the upcoming study. Additionally, the social understanding tasks did not correlate substantially with the Big Five personality traits. Altogether, results pointed towards the divergent construct validity of social intelligence with academic intelligence and personality traits. However, results also revealed some meaningful problems relevant to test modifications and extensions in Study 2.

Concept and Design of the Social Understanding Tasks

The intended faceted design of the social understanding tasks has not yet been accomplished. The present study was unable to provide convincing results regarding which item format worked best. This was because only the ratings-based scales included sufficient item numbers to investigate separate scales. What can be concluded from the present study is that the ratings-based scales seems to work as well or better than the remaining two item formats. The within-domain correlations also suggest the use of ratings-based scales because they showed a clearer pattern of intercorrelations compared to those based on all item formats.

Additionally, open response formats were seen to be uneconomical in terms of scoring. Therefore, it was decided to focus only on ratings-based formats in the consecutive task modifications.

The restriction to the use of only ratings-based scales in the final social understanding scales of the present study, however, posed another problem not directly visible from the results and not attended to prior to and during scale construction: The distribution of the target's answers was not balanced across the 6-point rating scale. Figure 7.13 presents the distributions of the target answers across the six possible rating categories, separately for the four targets *(a-d)* and overall targets *(e)*. Rating category "1", in particular, was overproportionally represented. This problem should be addressed at this point in order to decidedly conclude what modifications were required for Study 2.

The possible consequences of the unequal distributions were possibly skewed distributions of the target scores and an enforcement of a bias related to the rating tendencies of the judges. This represents an important problem for the present study. At this point in the present work, a more detailed analysis of the effect of this problem on the psychometric item properties and the validity results was not undertaken because of several reasons. Most importantly, it had already been difficult to select items for the final scales that approximated the intended taxonomic principles (because of an unbalanced conceptualization of the social understanding tasks). Any further item selection based on the distribution of the target answer would enhance the imbalance of item contents making the derivation of conclusions from this analysis difficult. Instead of only post-hoc statistical means in the present study, this problem should be addressed in Study 2 by attempting to imply more balance into the distribution of the target answers. Further scoring or statistical means will additionally be addressed in Study 2 (see Chapter 6.3 for the exploratory questions surrounding this problem).

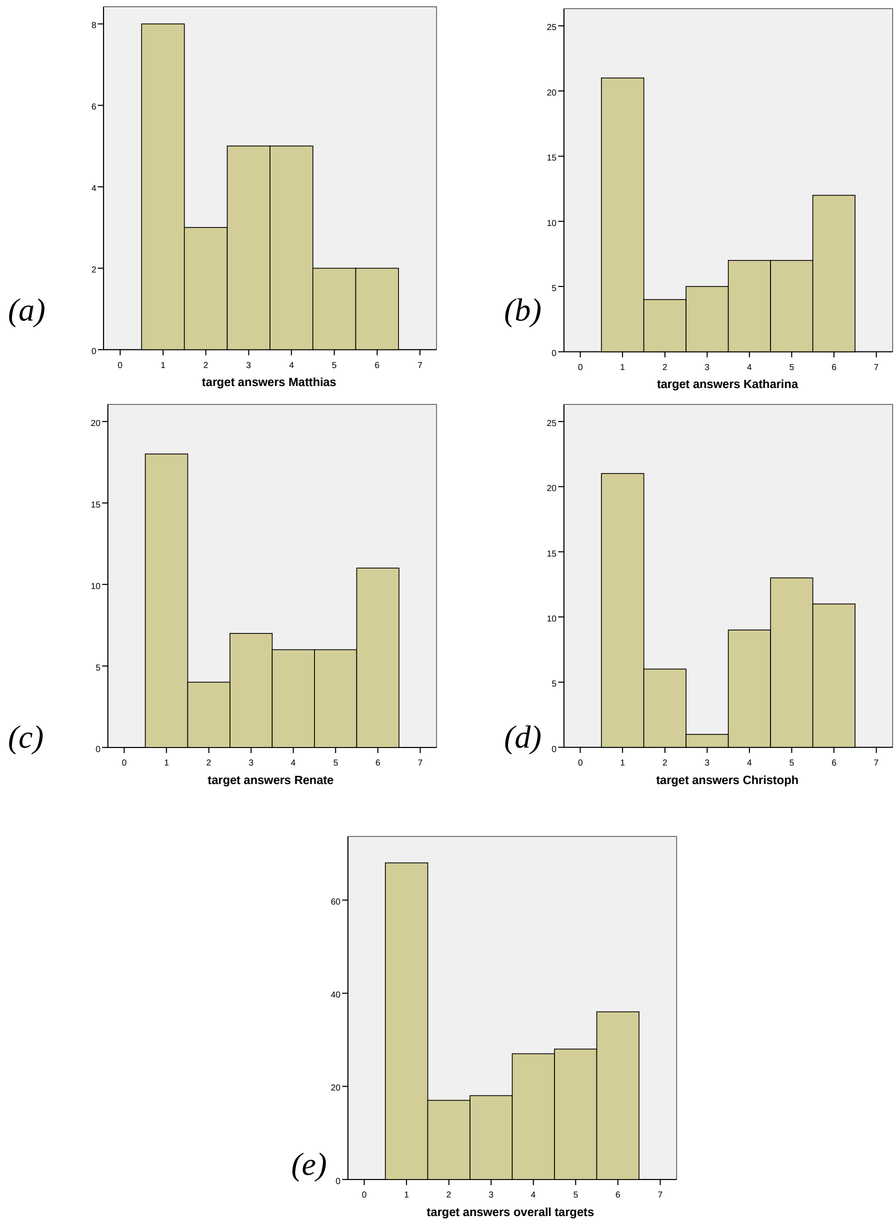


Figure 7.13

Distribution of the Target Answers on Ratings-Based Items in Scenarios (Study 1)

Note. Targets *(a)* Matthias, *(b)* Katharina, *(c)* Renate, *(d)* Christoph, *(e)* overall

The different modalities were also not equally represented in the items of social understanding tasks. Due to the large numbers of items excluded during item selection, the content-related scales did not include equal proportions of every target and every modality. Thus, the generalizability and representativeness was clearly reduced in the present study.

Specifically, it appeared difficult to assess the target's cognitions by the use of ratings-based scales. This modality was most frequently represented in open response formats. As a post-hoc explanation, the formulation of ratings-based items seems to have been difficult or at least counter-intuitive during the item construction. Subsequent test modifications needed to find a way to deal with this problem because only ratings-based scales were applied hereafter.

Number of Missing Values and Specific Psychometric Weaknesses

The screening of the data revealed problems with missing values, particularly in one pictorial and the video-based task of social memory. This observation conformed with the experiences during data sampling when subjects had complained of not having enough time to view the stimuli and answer the questions. Consequently, the resulting reliabilities were rather low which was attributed to the relatively small number of items. One of the tasks completely failed because of too many missing values. Subsequent test modifications are needed to account for this problem. Another concern related to the social memory tasks referred to *(a)* the transformation of picture- and video-based stimuli into adequate item formulations and *(b)* the scoring of the circumscriptions of these stimuli within the subjects' answers. It was decided that the items would be reformulated by using pictures as additional information and by including multiple choice format items based on pictorial stimuli as well. This should allow the presentation of pictures within the items in order to avoid circumscriptions as far as possible. Despite these problems within the social memory domain, it seemed surprising that the two tasks still fit into a structural model of social intelligence. Any problems encountered during confirmatory factor analysis could be attributed to a lack of reliability of these two tasks.

The social perception tasks did not show problems with the measurement quality. Validity results, however, did not support a consistent ability domain. Task intercorrelations were marginal except for the pictorial and the video-based task. These tasks, however, represented very similar requirements so that this result did not seem surprising. Attempting to provide a post-hoc explanation, the lack of convergent validity within the social perception domain could be attributed to disparate levels of "social complexity" or diverse targets referred to in the different content-domains. The auditory task required the identification of cues about interactive behavior during spoken conversations whereas the pictorial and video-based tasks "only" required the identification of target persons wearing different clothing. Therefore, the test modifications and extensions should focus on adding other types of targets in all four content domains to assimilate the requirements.

The main focus of the present study lay on the test development and thus, on the investigation of adequate item and response formats, scoring methods, and presentation and answering times. However, data also allowed the investigation of the construct validity in terms of the structure of social intelligence and its relationship to academic intelligence and personality traits (i..e., divergent construct validity). Confirmatory factor analysis supported a two-factor structural model of social intelligence (r = .35 between social understanding and social memory) and also a hierarchical solution with a higher-order general social intelligence factor. However, some measurement problems were encountered in the hierarchical model so that this model needed replication in Study 2.

The present study also supported the divergent construct validity of the social intelligence operative ability domains. The social memory factor, however, showed large correlations to the BIS operative domains BIS-Memory and –Reasoning so that it was not clear whether this factor was truly independent from academic intelligence. A last model postulated in confirmatory factor analysis, however, could prove the independency of the social intelligence structure from the BIS operative ability structure.

8 Study 2

8.1 Sample

A total of 190 subjects participated in Study 2. They received full monetary compensation (see Table 6.6) or half of the money and, in turn, detailed feedback about their results in the study. As requested by the "Deutsche Forschungsgemeinschaft" (DFG), an unselected sample was applied. Thus, subjects were recruited by a promotion booth on the Christmas market in Magdeburg 2005, flyers in medical practices and in public institutions, posters, and an editorial article in a daily newspaper in Magdeburg ("Volksstimme"). Only subjects between 23 and 40 years of age were recruited for the study. One subject was excluded from further testing after the first day because it turned out that she was only 19 years old. Three more subjects did not return for the second testing day without any obvious reasons. They were excluded from any further analysis. Three other participants demonstrated low compliance and comprehension problems during data sampling. An inspection of their written responses in several tasks strongly supported this impression and they were also excluded from the data set. One other participant was excluded while the answers to the paper-and-pencil tests were transferred into SPSS data files. The participant had shown extreme response tendencies in every task based on rating scales and always marked the lowest or highest rating category.

The final sample consisted of 182 subjects, heterogeneous in age, education, and occupations. Mean age was m = 28.69 (sd = 5.57), 107 were females (58.8 %). However, Figure 7.1 clearly shows that age groups were not equally represented in the sample. One hundred and forty had a German high school degree ("Gymnasium") or a corresponding degree which equaled 76.9 %, 37 of these had finished a university degree. Eleven subjects had a 12th grade high school degree ("Fachhochschulreife"), and 31 had finished German middle school ("Realschule"). All presently lived in the region around Magdeburg. The professions of the subjects ranged across diverse occupational fields, from academic professions (e.g., medical doctors, teachers, economists, graduate engineers, computer scientists etc.), service and nursing occupations, to commercial or mechanical occupations.

Appendix D presents the standard scores of the samples in Study 1 and 2 based on the normative sample of the academic intelligence test applied in both studies (BIS-Test; see Jäger et al., 1997). The aggregated scores for the cells, scales of the ability domains, and the general academic intelligence level are shown. The comparability of the scores was restricted since the tasks entering the aggregation were not always the same. However, most of the tasks were applied in both studies. Appendix D shows that the performance of the sample in Study 2 was slightly lower for some ability domains. The variance maintained at the same level. Consequently, the upcoming analysis needed to consider possible sample effects when the difficulty of the newly constructed tasks were interpreted.

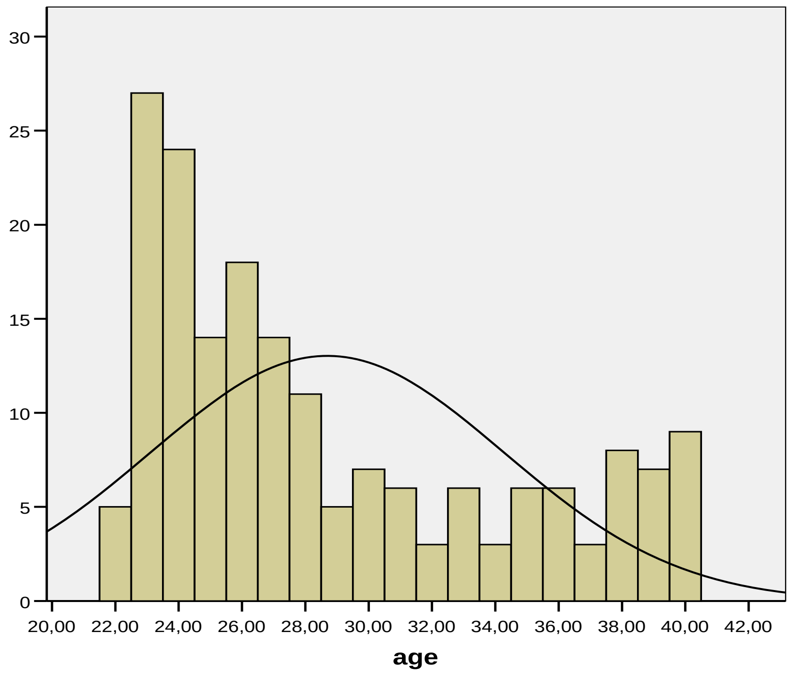


Figure 8.1

Age Distribution of the Sample in (N=182) (Study 2)

One possible way to deal with this sample effect was to correct for age effects that might have been responsible for the differences in the level of performance. Appendix D also presents the age correlations for the BIS cells, the scales of the ability domains, and the general academic intelligence level for the present sample. Several meaningful age correlations were discovered, particularly for the cells of the BIS-Reasoning domain (*r* between -.112 and -.316) and the BIS-Memory domain (*r* between -.153 and -.166). Surprisingly, the smallest correlations were found for the cells of the BIS-Speed domain where usually the largest correlations could have been expected (Cattell, 1971, 1987). Appendix D also presents the correlations with age after the academic intelligence scores were standardized for the respective age group based on the sample of Brunner and Süß (2005; N = 1247). The age groups were determined relying on sensible categories that were sufficiently represented in the reference sample. The age categories were between 23 and 25, 26 and 30, 31 and 35, 36 and 40. However, the correlations of the scores corrected for age showed a bias in the opposed direction. Now, the BIS-Speed cells showed positive age correlations (*r* between .064 and .204), the remaining correlations lay around zero with one more significantly positive correlation with the numeric reasoning cell ($r = .164$). Since the results of the age standardization were such equivocal, the uncorrected scores were applied hereafter. The specificity of the sample, however, needed to be accounted for in interpretation of the research results.

8.2 Material

8.2.1 Principles and Aims of Task Modifications and Extensions

The responsibilities for the task modifications and extensions remained the same as those applied in Study 1. The ideas and principles underlying the modifications, their realizations, and the additional construction of tasks based on written and spoken language can be referred to in the work of Seidel (2007). The present work focused on the pictorial and video-based tasks of social memory and social perception. The scenario presentation is included in both works since the test approach did not allow a material-related separation.

Social Understanding

Tasks of social understanding did not show psychometric problems related to the reliabilities of the final scales. Nevertheless, Study 1 suggested some substantial changes in the already existing scenarios and associated guidelines for the construction of new scenarios. The changes addressed, above all, some formal but also some content-related characteristics:

a) Technical Implementation:

In Study 1, the test length of the scenarios had varied largely which extended the planned total testing time substantially. Tasks were now implemented in Wmc 0.18. This circumvented the problem of subjects watching, listening to or reading stimuli for as long as or as often as they wanted to. The program did not allow returning to a previous page to start the presentation again, as was allowed by the PowerPoint presentation software in Study 1. Presentation time for pictures was limited to 10 seconds per picture, reading times were also limited to a comfortable but not extensive level (between one and three minutes for texts presented on maximally two screens). In general, these measures reduced testing time.

The implementation in Wmc 0.18 implied that video-based stimuli could no longer include an audio stream so that the self-presentations at the beginning of each scenario had to be shown via beamer on a large screen and loudspeakers (see Chapter 6.2.4 for the technical details). Responses were still given in paper-and pencil format in Study 2.

b) Instructions to scenes:

Another change related to the introduction to each scene. In the first study, subjects could first read the questions, then view or listen to the stimuli, and then read the questions again. The introduction was now extended by summarizing the upcoming questions in one or two sentences (see Chapter 6.2.2.1 for an example). It was expected that this change would additionally reduce the overall testing time.

c) Accomplishment of taxonomy:

To account for the taxonomic demands, the cross-classification of settings, modalities, and task material was accomplished within and across each scenario and resulted in a 8 x 4 x 2 x 3 design: eight targets, four types of task material, two types of settings (i.e., private vs. public), and three types of queried modalities (i.e., emotions, cognitions, and relationships).

d) Adjustment of item format

The item formats were changed to only 7-point rating scales. For items already based on rating scales, this basically meant a transformation of the visual analogous scale of the original target answers into seven instead of six categories. The most prevalent reason for the adjustment was that some rating dimensions demanded a true middle category (i.e., bipolar scales; for example, a target described himself as active vs. passive in relation to a third person). Since only one item format was to be applied throughout the scenarios, the unipolar rating dimensions and the personality ratings at the end of each scenario were also transformed into 7-point rating scales.

During the course of test modifications, the modalities of emotions and relationships could easily be queried by already existing items. Items that queried the cognition of the target were, so far, only in free response or multiple choice format. The change to rating scales required the construction of a new item type not included in the scenarios of Study 1. Therefore, the open-ended answers of the targets were translated into item formulations (e.g., one open-ended answer of Matthias was *"It's getting ridiculous, it's almost funny how she is constructing her thoughts, I'm angry that she doesn't stop misrepresenting my motives"*, the resulting item was *"What does Matthias think in this situation? How much would he agree with the following thoughts? (a) She doesn't understand me. (b) She intends to discriminate against me. (c) We are at cross-purposes."*). Sometimes, the targets had to be contacted again in order to get their answers to these newly formulated items on the visual analogous scale as the basis for the rating scales. This presented a problem because of the large time delay between the recording of the scene and the querying of the target information. However, the problem could not be solved otherwise.

The personality ratings were originally 5-point rating scales from 0 to 4. In order to determine the target score, the "correct" response receiving a score of "0"-deviation was calculated by a linear transformation of the 5- to a 7-point rating scale (i.e., from 0-4 to 1-7) by multiplying by 1.4 and adding 1.

e) Control questions were added at the end of each scenario (see Chapter 6.2.2.1).

f) Adding of four scenarios

Between Study 1 and 2, four scenarios were added enhancing the heterogeneity of the targets, the task material, and the situative contexts. This allowed the investigation of gender effects of judge and target for a broader target range.

g) Example scenario

An example scenario preceded the eight test scenarios. It was directed at making subjects familiar with the test principle and navigation between the scenes within the experimental program.

Social Memory

The social memory tasks mainly suffered from low reliabilities (SMp and SMf) and a large number of missing values. Moreover, the testing time devoted to the social memory tasks based on videos and pictures was restricted in Study 1.

a) Technical implementation: Social memory tasks except for those based on written language were now implemented in the Wmc 0.18 experimental program. This allowed fixed presentation times, and also fixed response time when answers were recorded by the PC (SMa2 and SMp1).

b) Enlarged number of items: More testing time was available for the social memory tasks. This allowed the enhancement of item numbers thus hopefully increasing reliability coefficients.

c) Enlarged presentation and answering time: The undesired large correlations with BIS-Reasoning and BIS-Speed which were even larger than those with BIS-Memory were assumed to be due to the speed effect included in the first task versions. Thus, presentation and response time were extended in order to reduce this effect. The extension of test time resulted in a lower number of missing values. At the same time, the amount and complexity of stimuli was reduced (e.g., less pictures in one sequence and shorter video scenes).

d) Presentation format: The presentation format of the picture sequences in the task SMp2 was changed. In the first version, subjects could see each picture for five seconds and could not return to an earlier presented picture. The task applied in the second study allowed deliberate browsing within one picture sequence and only an overall presentation time limitation was introduced. Thus, subjects could decide for themselves how much time they wanted to spend looking at one picture. This modification was also supposed to enhance real-life fidelity.

e) A new task was added based on auditory material. It attempted to find an equivalent task to the *Memory of Couples* task (SMp1). The task will be described in the upcoming material section.

Social Perception

The core problems related to the social perception tasks represented the lack of convergence between the tasks based on different task material. This was attributed to the different levels of complexity contained in the tasks. So far, only one task per cell was developed. Thus, the changes in the test approach centered on developing an additional second task per cell.

a) Task extensions: New task developments addressing the pictorial and video-based task concerned the extension of complexity of the task material and of the cognitive requirements. The complexity mainly referred to the types of target cues which represented social interactive cues instead of persons only in the two new tasks.

b) In turn, the new tasks based on written and spoken language were sought to be less complex.

c) The existing tasks had proven useful and did not show psychometric weaknesses. Therefore, the number of items and thus, the testing time was reduced. Some additional changes within these tasks concerned specific modifications of procedures and implementations. These will be addressed in the upcoming task descriptions.

8.2.2 Social Intelligence Tasks

Social Understanding Tasks

The test principle of social understanding tasks were already made familiar (see Chapter 6.2.2.1). In short, the scenario tasks required subjects to judge the emotions, cognitions, relationships, and the personality of target persons on 7-point rating scales. To operationalize every content domain, judgments were based on information from written and spoken language, pictures, and videos. In Study 2, eight scenarios were applied involving target persons Renate (SU_RF), Bringfried (SU_BS), Conny (SU_CK), Christoph (SU_CP), Katharina (SU_KL), Friedrich (SU_FB), Hannah (SU_HR), and Matthias (SU_MM) (see Table 6.2 and Figures 6.1 and 6.2 for the demographic and personality characteristics of the target persons). Subjects had to judge the targets' mental states in terms of three different modalities (i.e., emotions, cognitions, relationships) based on four different types of task material (i.e., written and spoken language, pictures, and videos) and two types of settings (i.e., private and public). Additionally, the personality traits of the targets based on nine dimensions (Big Five and Interpersonal Problem Circumplex) had to be judged. All items were 7-point rating scales. Table 8.1 presents the number of items in the faceted design of the social understanding tasks cross-classifying eight targets, three modalities, two settings, and four types of task material. Targets are numbered in the order of testing. Items were selected to equally represent taxonomic categories across all scenarios.

A total number of 115, 125, 113, and 124 items were applied for the scales related to the task material (for v, a, p, and f, respectively). No cell in the design stayed empty. Each scenario was planned to last about 20 minutes. There was still variance observed during data sampling. The average deviation from the planned task time was, however, substantially smaller than in Study 1. Cronbach's alpha coefficients for the final scales in the present sample were .77, .84, .79, and .83 for the scales SUv, SUa, SUp, and SUf.

Table 8.1

Number of Items in the Faceted Design of the Social Understanding Tasks (Study 2)

		Targets in order of testing*																		
		1 (RF)		2 (BS)		3 (CK)		4 (CP)		5 (KL)		6 (FB)		7 (HR)		8 (MM)				
		pr	pu	pr	pu	pr	pu	pr	pu	pr	pu	pr	pu	pr	pu	pr	pu	∑	∑	∑
	E	2	4	3	4	5	5	3	3	4	5	1	3	4	1	4	3	54		E_{tot}: 214
V	**C**	4	4	1	1	4	3	1	1	1	3	3	4	4	2	2	2	40	115	C_{tot}: 160
	R	1	1	1	1	1	1	1	1	1	1	1	1	3	1	4	1	21		R_{tot}: 130
	E	1	5	5	3	5	5	4	1	5	5	4	6	6	4	4	1	64		
A	**C**	5	5	4	3	3	2	4	2	1	4	1	1	3	4	3	3	48	152	
	R	4	1	1	3	1	1	4	2	4	4	4	4	1	1	4	1	40		
	E	5	1	4	5	3	4	1	3	4	1	4	1	1	5	1	5	48		
P	**C**	2	3	3	3	1	1	1	1	4	4	1	2	3	3	1	3	36	113	
	R	1	1	4	4	1	1	4	1	3	1	2	1	1	1	2	1	29		
	E	1	4	1	1	3	6	4	3	1	6	1	1	4	5	2	5	48		
F	**C**	3	1	4	3	1	1	4	1	3	3	2	1	1	4	2	2	36	124	
	R	4	1	1	1	4	4	1	4	4	2	1	1	1	3	4	4	40		
	∑	64		64		66		55		74		51		66		64				

Note. v = written language, a = spoken language, p = pictures, f = videos
e = emotions, c = cognitions, r = relationships
pr = private, pu = public

In contrast to the social understanding tasks in Study 1, the test development in Study 2 accounted for the positions of the target answers on the rating scale. It was attempted to select items that allowed an equal distribution across the rating categories. Figure 8.2 presents the distribution of target answers across the items for each of the eight scenarios and across all scenarios. Diverse distributions were achieved not implying as one-sided distributions as in Study 1. Equal distributions, however, were not achieved and only the distribution across all scenarios was balanced with a slight overrepresentation of the categories “1” and “5”. This problem should additionally be addressed by statistical means in the exploratory research questions at the end of the present study.

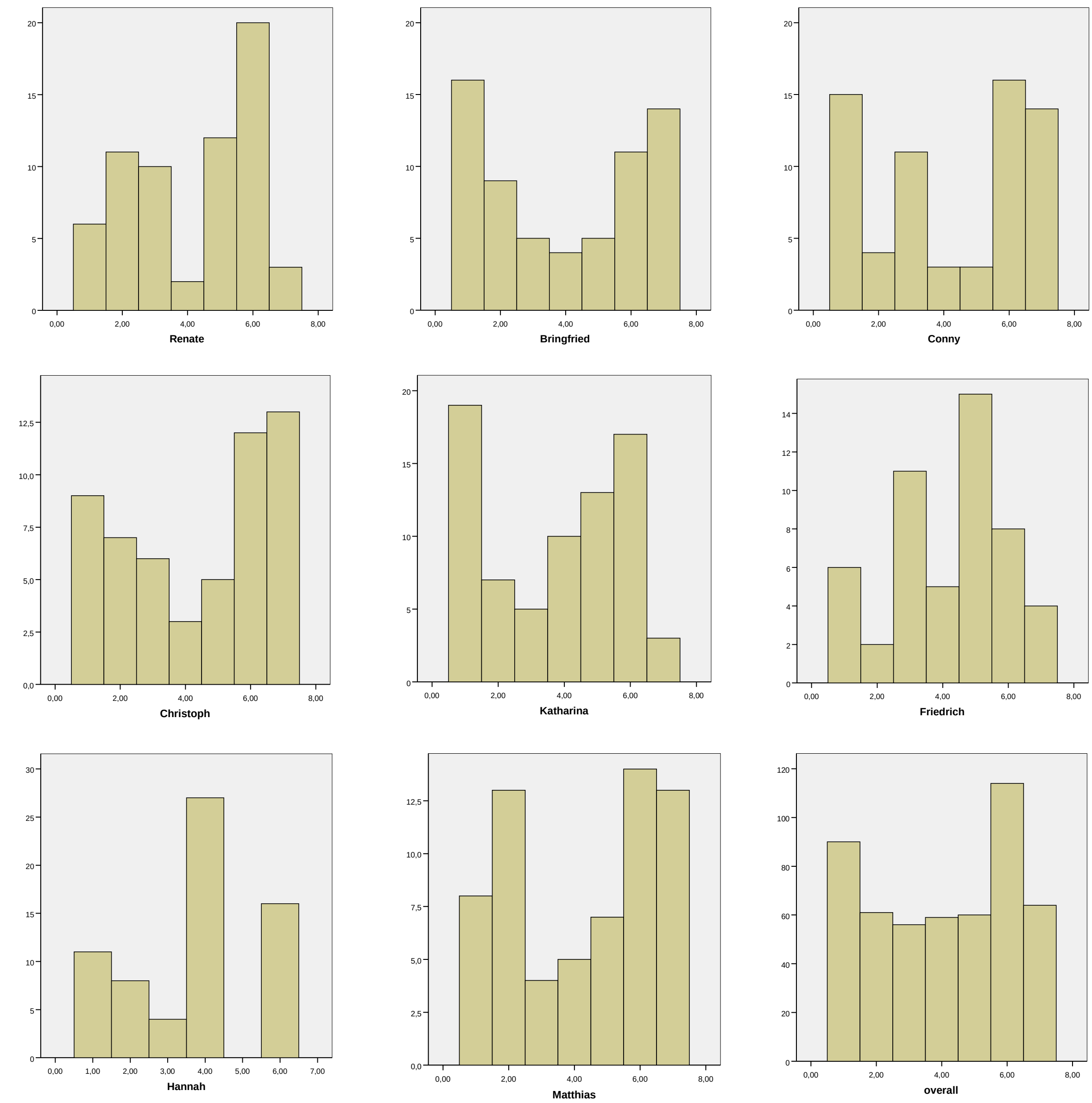


Figure 8.2

Distribution of Target Answers on Items in the Scenarios (Study 2)

Social Memory Tasks

Social Memory – written language (SMv): Memory for Written Correspondence

Participants were presented with six written one-way or two-way correspondences (e.g., a letter written after a skiing vacations) and told to read and to memorize as many socially relevant details as possible. Reading times varied between 1:30 min. and 2:30 min. Participants were then asked questions about these details . Response format was both, open-ended and multiple choice. In the open-ended response format items, subjects had to freely reproduce the information from the text (e.g., "What does the writer say that she felt sorry about?"). In the multiple choice items, subjects had to identify the correct answer out of five choice alternatives (e.g., “Which of the following reasons does the writer name for writing the letter instead of a postcard?”). Response time was limited (from 1:20 min. and 2: 20 min.). The test contained a total number of 48 items. The test was subdivided into two halves for data sampling since the full version was too long to maintain adequate concentration. The task showed a reliability of .84 (Cronbach’s alpha) in the present sample.

Social Memory – spoken language (SMa1): Memory for Conversations

Participants listened to six monologues or conversations between two or more people (e.g., a male person talks about a conflict at work with colleagues and superiors). Subjects had to memorize as many socially relevant details as possible. The length of the stimuli varied from 0:43 min. to 1:30 min. Participants had to answer 36 questions about these details in multiple choice and open-ended formats (e.g., “For what, according to the speaker, was he criticized at work?”). Response time was limited (from 0:45 min. to 1:45 min.). Cronbach’s alpha for the present sample was .80.

Social Memory – spoken language (SMa2): Memory for Voices

Participants listened to 12 conversations between at least two people. After each conversation, subjects were presented five different voices and they had to decide by the way of a mouse click on the respective number which of the five voices was heard in the conversation. Answers were scored in terms of the proportion of correct answers. The reliability (Cronbach’s alpha) for the present sample was .19.

Social Memory – pictorial 1 (SMp1): Memory for Couples (Weis & Süß, 2007)

This task required subjects to view and memorize pairs of people (i.e., heterosexual couples and colleagues of the same sex). In the recall section, subjects were shown one person and they had to identify that person’s corresponding partner from four alternatives. All the people portrayed in the recall section wore different clothing. The task’s reliability in the first study was not sufficient (i.e., $\alpha = .524$). For Study 2, the number of items was extended from 16 to 18. The former blocks consisting of eight pairs each, separated into either heterosexual couples (block 1) and colleagues of the same sex (block 2), were dissolved. In the present task, three blocks of five, six, and seven couples were presented. In each block, both types of pairs were presented (Block 1: 3/2; Block 2: 3/3, Block 3: 3/4, numbers representing couples / colleagues). Consequently, the choice of alternatives for one person contained partners from the opposite and the same gender. The recall section again followed directly after one presentation block. Each picture stayed on the screen for three seconds. Response time was limited to ten seconds. The task was implemented in Wmc 0.18, including the presentation and recall section. Subjects responded by way of a mouse click on the number of the correct choice alternative. Performance was scored as the proportion of correct answers. The reliability of the task in the present sample for the total of 18 items was .56.

Social Memory – pictorial 2 (SMp2): Memory for Situations - Pictures

Participants had to memorize socially relevant details from a sequence of pictures showing different numbers of people interacting within one situative context. The task contained one example and four test sequences with varying numbers of pictures. Table 8.2 presents the picture sequences in the order of testing, example items, the total number of pictures and items per sequence, and the presentation and response times. The task was implemented in the experimental software. This allowed some important changes. Presentation times were no longer fixed to one picture but to one sequence (see Table 8.2). Both, presentation and response times were extended. The response format was changed to multiple choice and open-ended formats with an emphasis on multiple choice items. With this, subjects did not have to freely reproduce item contents which was thought to have extended the influence of the speed component. Moreover, more objective and economic testing and scoring was allowed. The choice alternatives frequently consisted of pictures so that the problem of circumscribing pictorial information by written language was avoided. Appendix E presents two example items with pictorial choice alternatives. The examples stem from the first picture sequence (i.e., family get-together). The task contained a total of 22 items. The multiple choice items were scored in terms of the proportion of correct answers, and the free responses were rated by two raters in terms of the accordance with the correct answer. Credit points were given for the number of achieved points in relation to the maximum number of points. Cronbach's alpha for the task in the present sample was .46.

Table 8.2

Overview of Picture Sequences of the Task SMp2 Including Number of Items, and Presentation and Answering Times (Study 2)

Sequence	Picture count	Example item	Item count	Presentation time	Answering time
1: Family get-together: a group of adults having dinner and doing a boat trip	3	"Put the following people in the correct order according to their positions at the dinner table from left to right."	4 (3/1)	1:00	1:00
2: Ice cream parlor: a young woman having icecream	4	"How many times does the woman look into the menu together with the serving staff?"	5 (3/2)	1:20	1:20
3: Party: many young people at a party	5	"Which of the following pairs of people had eye contact?"	6 (5/1)	1:40	1:50
4: Staff room of teachers: teachers during their break	5	"Who of the following persons was seen smiling?"	7 (5/2)	1:40	2:00

Note. * in parentheses are the numbers of multiple choice / free response format items, respectively

Social Memory – video-based 1+2 (SMf; Weis & Süß, 2007): Memory for Situations - Videos

The tasks SMf1 and 2 only differed in the complexity of task material. No separate analyses, however, were conducted for the two sub-tasks and they will be treated as one (called SMf hereafter). The task comprised one example video scene and five test video scenes presenting different social events. Participants viewed each video scene once and were instructed to memorize as many socially relevant details as possible. After viewing each scene, subjects had to answer open-ended and multiple choice items about details. The item construction for Study 2 included the same principle that was just described for the pictorial social memory task SMp2 (i.e., pictorial choice alternatives). It was intended to achieve a smaller number of missing values by using shorter video scenes and providing longer response times. Table 8.3 presents an overview of the five scenes including a short description, an example item, the numbers of items, the stimuli length, and the answering times.

Table 8.3

Overview of Video Scenes of the Task SMf1+2 Including the Number of Items, and Presentation and Answering Times (Study 2)

Sequence	Video length*	Example item	Item count**	Answering time
1: In a pub: several male adults are seen in a pub, talking and playing cards	1:13	"Which of the following gestures was included in the scene?"	7 (5/2)	2:00
2: Group work: three young people are working together at one table	1:30	"What does the young woman do when the young male leaves the room?"	7 (4/3)	2:00
3: A bowling night: a group of seven adults during a bowling night is seen sitting together at one table.	1:07	"How does the woman returning to the group learn what has happened while she was away?" (Multiple Choice)	7 (4/3)	2:00
4: Staff room of teachers: teachers during their break involved in different activities	1:32	"Which is the correct order of appearance of the following situations depicted in pictures?"	7 (3/4)	2:00
5: Snooker training: a coach is seen giving advice and several others are practicing at the tables	1:36	"With whom of the following persons does the coach have a Einzelbetreuung?"	7 (5/2)	2:00

Note. * in minutes and seconds
** in parentheses are the numbers of multiple choice / free response format items, respectively

Multiple choice items were scored in terms of the proportion of correct answers, and free responses were rated by two raters in terms of the accordance with the correct answer. Credits were given for the number of achieved points in relation to the number of available points. The reliability of the task in the present sample was .65. Finally, Table 8.4 presents the taxonomic classifications achieved by the social memory tasks based on pictures and videos.

Table 8.4

Taxonomic Principles Underlying the Social Memory Tasks Based on Pictures and Videos (Study 2)

		Setting	
		Private	**Public**
	Two people	SMp1: heterosexual couples	SMp1: colleagues
Number of persons involved	**Small groups (number of persons in parentheses)**	SMp2 sequence 1: family excursion (5) SMf scene 1: pub (6) SMf scene 3: bowling night (6) SMp2 sequence 3: party (many)	SMf scene 2: group work (3) SMp2 sequence 2: ice cream parlor (5) SMf scene 5: snooker training (9) SMp2 sequence 4: teacher (16) SMf scene 4: staff room teacher (many)

Social Perception Tasks

Social Perception – written language 1 (SPv1): Perception of Social Cues - Texts

Subjects were presented with either one or two written target statements or questions on one half of the screen (e.g., "Does the sender criticize a lack of engagement?"). Based upon a subsequent short text (e.g., "I think you're engagement is great.") presented on the other half of the screen, they had to decide as quickly as possible about the truthfulness of the statements ("true" vs. "false") or whether the question could be answered with a "yes" or "no" response (i.e., choice reaction time). When two statements or questions were applied, both needed to be true for a "true"-answer. Target statements or questions and the text could be seen simultaneously for some time. Subjects indicated their decision by pressing the keys representing their choice. Response time was limited to 20 seconds . Performance was scored in terms of the reaction time for correct responses also accounting for false alarms. The task consisted of 60 items, arranged in two blocks with 30 items each. The first block contained items with one statement or question, the second block items with two questions or statements. Subjects had the possibility to take a self-timed break in the middle of each block. Cronbach's alpha reliability coefficient for the present sample was .97.

Social Perception – written language 2 (SPv2): Perception of Social Cues - Texts

This task was newly developed and intended to contain less complex reading requirements than the task SPv1. The task consisted of short sentences presented in two blocks with 30 sentences each. In the first block, subjects had to decide whether the sentence expressed socially relevant contents or not (e.g., social: “Don’t feel burdened by a bad conscience.”, not social: “The accounting is only preliminary.”). In the second block, subjects had to decide whether the sentence expressed something positive or negative (e.g., positive: “I’d be happy to keep contact with you.”, negative: “I don’t feel really well.”). Both tasks represented choice reaction time tasks. Within one block, subjects could take a self-timed break. Care was taken to ensure that each item could be answered unequivocally. Answers were scored in terms of the reaction time for correct responses also accounting for possible false alarms. Cronbach’s alpha in the present sample was .98.

Social Perception – spoken language 1 (SPa1): Perception of Social Cues in Spoken Language

Subjects listened to seven extracts from conversations (e.g., a conversation about a birthday present). Prior to presentation, subjects were instructed to attend to specific target cues within the conversation. Examples of target cues were provided. Targets differed across conversations and varied in complexity (e.g., mentioning of a name, filling words, agreement or disagreement, change of speaker, etc.). Subjects had to react as quickly and as accurately as possible by pressing the key assigned to the target cue (e.g., “<” when agreement is expressed, “-“ when disagreement is expressed). Either one or more than one cue had to be attended to (simple vs. choice reaction time). Response time was limited in terms of the time frame until the next cue emerged within the recordings. Performance was scored in terms of the reaction time for correct responses also accounting for false alarms. The task consisted of a total of 93 items. Cronbach’s alpha for the present sample was .91.

Social Perception – spoken language 2 (SPa2): Perception of Emotions in Voices

Subjects were required to decide which of two given emotions was expressed in a spoken statement (e.g., irony vs. anger, positive vs. negative, etc.). The task consisted of 6 blocks, each block containing ten sentences. In the first three blocks, stimuli contained real and sensible sentences. In the next three blocks, stimuli consisted of senseless sentences only spoken with a certain prosody. Subjects were instructed not to attend to the content of the spoken real sentences. Performance was scored in terms of the reaction time for correct responses also accounting for false alarms. The reliability (Cronbach’s alpha) in the present sample was .92.

Social Perception – pictorial 1 (SPp1): Person Perception – Pictures

This task required subjects to detect a given target person within pictures of crowds at public locations (e.g., a market place, a store, a mall, a pedestrian zone, etc.). Prior to the trials, the target person was presented in portrayals showing the person’s whole body. The targets wore different clothing in the prior presentation and the test trials so that subjects had to attend to the person itself and not to, for example, a specific jacket. Targets had to indicate as quickly as possible the location of the target person by a mouse click on the its head. The position of the targets in the pictures varied unsystematically. Contrary to the task in Study 1, only one target had to be attended to within one block of pictures. The items for the present version were selected so that the target was equally visible in the pictures in order to avoid missing values. The task comprised one practice block of five pictures and three test blocks including 15 trials each. Response time was limited to ten seconds. Additionally, unlike in Study 1, subjects had to execute a complementary click on a dot on the screen prior to each single test trial. This was intended to control the mouse position before each trial to provide each subject with identical conditions prior to each test trial. Performance was scored in terms of the mean reaction time of correct trials. Cronbach’s alpha in the present sample was .91.

Social Perception – pictorial 2 (SPp2): Perception of Body Language – Pictures

The task required subjects to decide which of two or three given target cues was presented in a picture. The target cues varied between blocks. The task consisted of one practice block and three test blocks. Examples prior to one block illustrated the target cues. Table 8.5 presents the different blocks and the respective cues, the item contents, and the number of items. Pictures from one situative context were spread across the blocks so that one block could not be classified into one specific taxonomic class. The situations involved between three and twelve persons. Subjects had to react as quickly and as accurately as possible by pressing the key assigned to the target cue (e.g., “<” when the picture showed a gesture independent from the conversation, “-“ when the picture showed a gesture illustrating the conversation). Response time was limited to ten seconds. Performance was scored in terms of the reaction time for correct responses. The task consisted of 43 items. Cronbach’s alpha for the present sample was .95.

Table 8.5

Target Cues of SPp2, Taxonomic Classifications and Item Numbers (Study 2)

Block	Target cues "Does the picture show ...	Item contents	Item count
Example	(1) A gesture independent from the conversation (2) A gesture helping to illustrate the conversation	Pictures across the sequences contained the following situations: - a barbecue - a meeting at work - group work - a wedding - a relocation - mechanics at work - a party	7
1	(1) A person looking at another person (without eye contact) (2) Two persons having eye contact		13
2	(1) One person smiling or laughing (2) Two persons smiling or laughing with each other		12
3	(1) One person watching someone else working (2) More than one person working on their own (3) More than one person working together		11

Social Perception – video-based 1 (SPf1): Person Perception – Videos

The task required subjects to detect one given target person within videos of crowds at different public locations (e.g., a market place, a store, a mall, a pedestrian zone, etc.). Prior to each block of test trials, the respective target person was portrayed in a short video extract showing the whole body. The targets wore different clothing in the presentation and in the final test trials. During the trials, subjects had to react as quickly as possible by pressing the space bar when they detected the target person. The task consisted of one example block with five videos and three test blocks with ten videos each. Contrary to Study 1, only one target had to be attended to within one block of trials. The items for the present version were selected such that the target was equally visible in the videos in order to avoid missing values. Response time was limited in terms of the length of the video. The length varied between 7 and 27 seconds. The target appeared in the videos at varying points in time, the point of emergence was determined in terms of the respective frame in the video. Target appearance across the videos ranged from the first frame to frame 479 so that subjects' expectations about the target appearance were not built up (one frame equaling 40 ms). Performance was scored in terms of the mean reaction time of correct trials accounting for false alarms prior to target appearance. The reliability coefficient in the present sample was .71.

Social Perception – Video 2 (SPf2): Perception of Body Language – Videos

The task required subjects to decide when one target cue of possible one, two or three given targets cues appeared in short video scenes. The target cues varied between the scenes. The task consisted of one practice scene and four test scenes. Examples prior to one scene exemplified the target cues. Table 8.6 presents an overview of the scenes showing the respective targets, the taxonomic classifications of the scenes and the number of items per scene. Subjects had to react as quickly and as accurately as possible by pressing the key assigned to the target cue (e.g., “<” when the picture showed a gesture independent from the conversation, “-“ when the picture showed a gesture illustrating the conversation). Response time was limited to terms of the time until the next target cue emerged in the video. Performance was scored in terms of the reaction time for correct responses accounting for false alarms. The task consisted of 37 items. Cronbach’s alpha for the present sample was .84.

Table 8.6

Target Cues of SPf2, Taxonomic Classifications and Item Numbers (Study 2)

Video scene	**Target cues** ”When does the video show ...	**Taxonomic classification**	**Item count**
Example	(1) A gesture independent from the conversation (2) A gesture helping to illustrate the conversation	A male person conducting a seminar (public, one person involved)	11
1	(1) A person showing joy	A woman a conversation on the telephone (private, one person involved)	6
2	(1) Someone turning away from someone (2) Someone turning towards from someone	A conflict conversation between a couple (private, two persons involved)	7
3	(1) A man starting the conversation without prior cue (2) A man reacting on someone by starting to talk	A male person conducting a seminar (public, one person involved)	8
4	(1) Both men focusing their attention on the other one (2) One or both focusing their attention on a task or another person	Two male persons having a meeting with others (who were not visible) (public, two persons involved)	5

8.2.3 Validation Instruments

Academic Intelligence: BIS Test (Jäger et al., 1997)

The BIS-Test served as the test for academic intelligence (see Chapter 7.2.2). Because of time constraints, the complete test could not be applied. The task selection was based on conceptual considerations. Tasks of BIS-Creativity were not included because an equivalent operational ability domain was not operationalized for social intelligence. Of the remaining nine cells of the BIS-Model, three tasks for each cell were selected that allowed a modeling of the BIS with equal contributions from the relevant ability domains. One task of the cell MV (Memory verbal; task “ST” = “sinnvoller Text”) was modified because it included social contents in the original version. The task was substituted by a text about “ground” (“Boden”) corresponding in terms of word count and number of items which did not include social contents. Additionally, two tasks were applied that represented supposedly easier versions of the numeric speed tasks “XG” and “SI" (XG = “x larger by ...”, “SI = divided by seven”). They were simplified by enclosing different numeric operations (e.g., “larger by 2” instead of “larger by 3”; or “divided by 5” instead of “divided by 7”). These tasks, however, did not substitute the original tasks but rather complement the test battery because of previously experienced problems with the task difficulties.

A total of 30 tasks including a warm-up task was partitioned into four parts of seven to eight tasks. The parts lasted about 20 to 25 minutes and were administrated at different points in time throughout the two testing days (see Appendix F for the order of testing including instruction and working time per task). Cronbach’s alpha reliability coefficients in the present sample for the 6 ability domains were .82 / .82 / .85 / .75 / .78 / .83 for BIS-S / -M / -R / -F / -V / -N based on parcels of each domain.

Personality: NEO-FFI (Borkenau & Ostendorf, 1993)

The NEO-FFI served as a measure for the Big Five *Neuroticism (NEO-N)*, *Agreeableness (NEO-A)*, *Extraversion (NEO-E)*, *Conscientiousness (NEO-C)*, and *Openness (NEO-O)* (see Chapter 7.2.2). Reliability coefficients (Cronbach’s alpha) in the present study were .87 (NEO-C), .76 (NEO-A), .69 (BIS-O), .79 (NEO-E), and .84 (NEO-N).

Profile of Nonverbal Sensitivity (PONS; Rosenthal et al., 1979) – Video

The Video subscale of the PONS was applied as a measure of nonverbal sensitivity. The full PONS could not be applied because all language-based stimuli were in the English language and not enough testing time was available. A test description is provided in Table 5.3 (for the psychometric properties) and in Appendix A (scale description). The original Video subscale contained 40 items (i.e., presenting either face or body). 20 items were added displaying face and body simultaneously. The original 40-item Video PONS did not show good reliability coefficients in previous studies (alpha < .40) and it was hoped to improve the reliability with the additional items. In the present study, however, the PONS-Video subscale showed a reliability coefficient of .30 (Cronbach’s alpha), and the extended 60-item version showed a reliability of .40.

8.2.4 Instruments Peripherally Related to Research Questions

Baseline Measures

The *Simple Reaction Time Task* (SRT; Oberauer et al., 2003; Sander, 2005) and the *Mouse Speed* task (MT; Oberauer et al., 2003; Sander, 2005) assessed as baseline measures for the social perception tasks were the same as in Study 1 (see Chapter 7.2.3). The reliability coefficients for the present sample showed equally high parameters (.96 and .98, for SRT and MT).

Readspeed

Seidel (2007) developed a new task to assess the readspeed baseline for the social perception tasks based on written language (SPv1 and SPv2). The task in Study 1 did not truly prove whether the subjects in fact read every single word. The new task was based on the Working Memory literature (Daneman & Carpenter, 1980). It included 60 sentences presenting either senseless or reasonable contents. Subjects had to decide for each sentence whether it was true or false and press the respective key on the keyboard as quickly as possible. Performance was scored in terms of reaction time for correct answers. Cronbach's alpha for this newly constructed task was .98.

Long Term Memory Task (LTM)

This task was not included in the core design of the SIM. However, long term memory could be conceived as a relevant domain of social abilities (Bless et al., 2004). It was assessed relying on the stimuli and background information contained in the social understanding tasks. Participants had not been instructed to memorize the information included in the scenarios. It was expected, however, that subjects were capable to remember this information because they processed it during working on the scenarios. The task was administered at the very end of the second testing day. The delay between the last scenario (Matthias) and the long term memory task was about 30 minutes. Thus, for the scenarios of the second testing day, the delay between stimuli presentation and recall was the same for every participant. For scenarios of the first testing day (Renate, Bringfried, Christoph, and Conny), the delay from stimulus presentation to the recall questions depended on the individual's dates of testing. The difference between the two testing days ranged from one to 32 with a mean difference of 7.34 days (*sd* = 5.36). 50.8 % of the participants completed the two testings within 7 days, another 16 % within just one day. The question of influence of this delay on performance however, is an empirical one and the delay should be controlled for if an effect is found.

The task consisted of seven items per scenario. Four items queried specific information contained in the task material, one item for each type of material (e.g., "What did Matthias say to his father as he helped him clear up?" ... *(a)* "You should have asked me before.", *(b)* "You helped me a lot.", etc.). One item presented an extraction from one of the scenes of one scenario and asked the participant to determine the correct target person. Two items referred to background information about the target persons given in the introduction or in the descriptions of the situations at the beginning of one scene. Item formats of the resulting 56 items were both, multiple choice (with five alternative) and free response. Performance was scored in terms of the proportion of correct answers for items based on multiple choice formats, and as the points achieved in relation to the available points for free responses. The latter were rated by two raters. Cronbach's alpha for the entire scale including 56 items was .82.

Self-Report Questionnaires

Exhaustion Questionnaire

At certain points during one testing day (two per day, see Procedures in the next Chapter), the participants' subjectively perceived exhaustion was assessed on the following dimensions - tiredness, exertion, and weariness. Responses were given on a 6-point rating scale from 1 (*very rested / strained / awake*) to 6 (*very tired / fresh / weary*), for each of the three dimensions. Item 2 was reversely coded.

Biographical Questionnaire Including Musical and Computer Experience (Feigenspan, 2005; Süß, 1996)

The biographical data sampled from each participant addressed age, gender, level of education, number and age of children if any, occupation, and mother tongue. Additionally, hearing capabilities (two choices: normal vs. disabled), vision, self-assessed musical abilities in relation to members of the same age group (5-point rating scale), and the type of musical experience, if reported was collected. The computer experience questionnaire separately assessed in Study 1 was embedded in the present questionnaire and shortened. Items addressed the time spent at the computer (i.e., multiple choice: several hours *(a)* per day *(b)* per week *(c)* per month and *(d)* less) and self-assessed computer experience in relation to members of the same age group (5-point rating scale). Finally, subjects had to split their time spent at the computer into the activities they are typically engaged in (in terms of percentage of the entire time spent at the computer). The following activities were included, *(a)* emailing, internet, computer games, *(b)* text and data processing, and *(c)* programming activities.

Social Behavior Questionnaire (Amelang et al., 1989)

The social behavior questionnaire based on prototypical acts was applied to assess self-reported social behavior (see Chapter 7.2.3). The reliability coefficient (Cronbach's alpha) for the present study was .88.

Schutte Emotional Intelligence Scale (SEIS; Schutte et al., 1998)

The SEIS was applied to assess self-reported emotional intelligence. Subjects had to indicate their agreement with 33 given statements. Answers were given on a 5-point rating scale from 1 ("*I completely disagree.*") to 5 ("*I completely agree.*"). Cronbach's alpha in the present sample was .85.

Empathy (Enzman, 1996)

The 14-item questionnaire of empathy was extracted from Enzmann (1996). Only the subscales; *empathic compassion* and *cognitive perspective taking* were selected, each represented by seven items. The questionnaire was translated by Enzmann from the English version of Davis (1980). Subjects had to indicate their agreement with the given statements. Answers were based on a 5-point rating scale from 1 ("*I completely disagree.*") to 5 ("*I completely agree*."). The two seven-item subscales showed reliability coefficients of .75 and .74 for perspective-taking and empathic compassion, respectively.

Altruism (Fahrenberg et al., 2001)

The 12-item Altruism scale of the FPI-R (Scale 2: social orientation; Fahrenberg et al., 2001) was applied. Subjects had to judge each of the 12 statements in terms of whether they agreed with it or not on a 2-point rating scale (1 = true, 2 = false). The reliability (Cronbach's alpha) in the present sample was .59.

Depression (Alter & Muff, 1979)

The Depression questionnaire was extracted from a list of social behavioral indicators of depressive symptoms that stemmed from a Delphi-study by Alter and Muff (1979). The questionnaire contained 24 items based on a 4-point rating scale from 1 ("*I do not agree*.") to 4 ("*I agree.*"). Cronbach's alpha in the present sample was .88.

8.2.5 Instruments Not Related to Research Questions

As in Study 1, several instruments not related to any research question or analysis were applied. They are listed as follows, including information about the intended measurement constructs. Detailed task descriptions and statistical analyses related to these tasks are described in Seidel (2007).

General Auditory Tasks

General Auditory Tasks – Nonverbal Tonal Tasks

- *Recognition of Repeated Tones* (Stankov & Horn, 1980): a 15-item performance task to assess the ability to recognize one tone within a sequence of eight tones that was played only once.
- *Tonal Series* (Stankov & Horn, 1980): a 15-item performance task that assessed the ability to identify the logical completion of a tonal series of four tones by selecting the correct tone out of four alternatives.

General Auditory Tasks – Nonverbal Tasks

- *Rhythm Reproduction* (Stankov & Horn, 1980): a 19-item performance task that required participants to reproduce rhythms by a keystroke on one key, rhythms could vary in length and complexity.

General Auditory Tasks – Language-Based Tasks

- *Masked Words* (Stankov & Horn, 1980): a 27-item performance task assessing the ability to identify spoken works (e.g., “tree”, “table”, etc.) against a noisy background varying in intensity (i..e., typical noise of a party), responses had to be written down in free format.
- *Audiobook*: a 19-item performance task requiring subjects to answer multiple choice items (with 5 choice alternatives) based on a previously memorized text, the text contained facts about a journey report to the Island Macao without including any socially relevant details.
- *Dissected Sentences* (Stankov & Horn, 1980): a 16-item performance task assessing the ability to rearrange previously disarranged words in order to compose a meaningful sentence, participants had to freely produce a sentence and write it down.

Working Memory Tasks

Tasks of Working Memory were sought to measure the simultaneous storage and processing as one of the core executive functions established in models of working memory (Oberauer et al., 2003; Sander, 2005; Süß et al., 2002). These were especially intended to validate the general auditory tasks in the study of Seidel (2007).

- *Memory Updating – numerical, adaptive Version* (MUN; Sander 2005): Subjects saw a 3x3 matrix on the screen, of which some cells were shaded. Only the non-shaded cells had to be attended to. Numbers appeared and disappeared in the non-shaded cells and had to be memorized by the subjects. Numerical operations (upward arrow indicating +1; downward arrow -1) appeared in the cells and had to be applied on the previously remembered numbers. The result of the operation had to be remembered again. After a series of arrows, a question mark appeared asking for the correct outcome in the respective cell. Difficulty was related to the number of non-shaded cells. The levels of difficulty in the adaptive version ranged from one to six non-shaded cells. When a level was accomplished, the next level started.
- *Word Span, adaptive version* (WS; Sander, 2005): Subjects were presented with a series of words on the screen (between three and nine words). Words had to be remembered and put into the order of the physical size of the object denominated by the words. The first letters of the words then had to be reproduced in the correct order. Difficulty levels were associated to the numbers of words presented. When a level was accomplished, the next level started.
- *Dot Span adaptive version* (DS; Süß et al., 2002): Subjects were presented a 10x10-matrix on the screen in which dots appeared and disappeared successively. The dot positions had to be remembered and reproduced. Additionally, after the last dot was presented, subjects had to judge whether the pattern was symmetrical (vertically, horizontally, or both) or asymetrical. After this judgment, subjects had to indicate by the way of a mouse click in the empty matrix where the dots had been positioned. Difficulty depended on the number of dots presented (between 2 and 6). When one level of difficulty was accomplished, the next level started.

Table 8.7 presents an overview of the design of Study 2 including only tasks related to any research questions.

Table 8.7

Design Presenting Performance Tests Related to the Research Questions (Study 2)

		Methods			
	Construct	**Written language (V)**	**Spoken language (A)**	**Pictorial (P)**	**Video-based (F)**
Social Intelligence	**Social Understanding (SU)**	Eight scenarios: SUv, SUa, SUp, and SUf			
	Social Memory (SM)	SMv1	SMa1	SMp1	SMf1
		SMv2	SMa2	SMp2	SMf2
	Social Perception (SP)	SPv1	SPa1	SPp1	SMf1
		SPv2	SPa2	SPp2	SMf2
		Written language (V)		figural-spatial (F)	numerical (N)
Academic Intelligence: BIS-Test	**Reasoning (R)**	WS		AN	RD
		WA		CH	ZN
		TM		AW	SC
	Memory (M)	WM		WE	ZZ
		ST („Boden“)		OG	ZP
		PS		FM	ZW
	Speed (S)	UW		ZS	SI (7/5)
		TG		BD	XG (3/2)
		KW		OE	RZ
Further Tests	**Social Understanding (SU)**				Profile of Nonverbal Sensitivity (PONS, Rosenthal et al., 1979)

Note. SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

8.3 Procedures

Chapter 6.2.4 included a detailed description of the procedures of both studies. For Study 2, Appendix F shows the order of administration and the planned testing time per task for testing day 1 and 2. Again, the order reflects changing requirements, concerning different task contents (e.g., spoken language vs. video-based tasks) and operations (e.g., reasoning requirements vs. self-reports, etc.). Testing was planned for about six hours, including four sessions and three breaks of five to 20 minutes each. Again, variance in the testing time per task was observed for tasks where participants were independently allowed to time their responses (i.e., the scenarios, the Working Memory tasks, and self-report questionnaires).

Since instructions and starting time for each task was coordinated for the entire testing group, there was about a half hour variance in total testing time per day.

After the first and the last session each day, the participants' subjectively perceived exhaustion was assessed on the dimensions of tiredness, exertion, and weariness. All three dimensions were highly intercorrelated within the four measurement points (r = .48 - .77) and were aggregated to compute a combined score of exhaustion. Table 8.8 presents the means and standard deviations of the four measurement points (T1 and 2 for Day 1 and 2, respectively), where, a high score is indicative of low exhaustion. The mean differences from T1 to T2 for both testing days were significant at .001 level (Day 1: t = 14.100, df = 180; Day 2: t = 11.757, df = 181).

Table 8.8

Means (m) and Standard Deviations (sd) of the Reported Exhaustion During Testing (Study 2)

	Day 1 T1	**Day 1 T2**	**Day 2 T1**	**Day 2 T2**
M*	3.745	2.730	3.873	2.995
SD	0.834	0.934	0.871	0.998

Note. * 1-6 rating scale, a high score indicating low exhaustion
T measurement point

Although the level exhaustion increased from the first to the last test session, it did not reach the low end of the scale which would have indicated extreme exhaustion, but stayed at the medium level. Whether exhaustion had an effect on task performance, will be statistically examined.

8.4 Results

8.4.1 Preparatory Data Analysis

The steps conducted prior to the main analyses in order to check the data for missing values and distribution problems relied on the same principles as described in Study 1. They shall be repeated briefly here.

1. Items or subjects were excluded when the number of missing values exceeded 15 % of the data points. Outlying data points were inspected and treated only if they influenced bivariate distributions or if an implausible deviation from the sample mean occurred.

Data treatment of reaction time scores (i.e., social perception and baseline measures):

2. Trials based on a wrong answer were set to missing.
3. Trials preceded by a false alarm were set to missing.
4. Reaction times lower than 100 ms were set to missing.
5. Based on the screening of the distribution of a first compound score, if necessary, reaction times of single trials slower or faster than 3 *sd* above the mean of the sample were set equal to 3 *sd* (trimming of outliers on the group level).
6. If necessary, reaction times of single trials of 3 *sd* above or below the mean of the individual were set equal to 3 *sd* above the individual mean (trimming on an individual level).

Social Understanding Tasks

The social understanding tasks in Study 1 had not exhibited meaningful problems related to missing values (i.e., four subjects were excluded because of a large number of missing data). In Study 2, neither any item nor subject exceeded the permitted number of missing values. No item had more than four missing values out of 182 subjects; 90 % of the subjects had three or less missing values out of 576 items and a maximum missing number of 19.

Social Memory – SMp1

No problems related to missing values had been encountered in this task in Study 1. In the present study, the task was only computer administrated including stimulus presentation and answers. Subjects were not forced to decide between the alternatives but could also wait until the next item emerged on the screen. No item had more than 15 % missing values. However, five subjects exceeded the threshold of 15 % missings (i.e., up to 14 for 18 items) and were excluded. For the remaining subjects, a missing value was set equal to a wrong answer.

Social Memory – SMp2

This task was not included in the final analyses in Study 1 because there were too many missing values. In the present study, the presentation and answering times were extended. Of the original 24 items in the present task, four items substantially exceeded the threshold of 15 % missings (i.e., up to 76 missings from 182 subjects). These items were always the last items within one picture sequence so that this result suggested that these items could not have been worked on within the available time allotted. Hence, these items were excluded from the analysis. Furthermore, five subjects had a large number of missings across the remaining items and were excluded from the analysis of this task. For the remaining items and subjects, a missing value was set equal to a wrong answer.

Social Memory – SMf

The same missing problem occurred for the video-based social memory task in Study 1 so that the answering times were extended and the videos were shortened. No item had more than 15 % missing values in the present study. Three subjects, however, had more missing values than allowed and were excluded from the dataset (i.e., up to 14 missing values for 40 items). For the remaining subjects, a missing value was set equal to a wrong answer.

Social Perception – SPp1

Missing values were counted prior to any recoding due to wrong answers. Therefore, a missing value represented the absence of a response. All items and subjects showed only a marginal number of missing values (i.e., up to four missings in 45 items; up to 11 missing data points in 182 subjects, respectively for subjects and items). After recoding the wrong answers into missing values, the number of missings in six items exceeded the acceptable level (i.e., between 34 and 93 for 182 subjects); these items were excluded from the analyses. The general level of correctness, however, was still high as was expected from Study 1. Figure 8.3 presents the distribution of wrong answers.

No reaction times were lower than 100 ms. A screening of the distribution of a first compound score of the remaining 39 items showed a slightly skewed distribution with several outlying values slower than 3 *sd* away from the mean. Therefore, reaction times of single trials slower than 3 *sd* were set equal to 3 *sd* (i.e., step 5). The distributions prior to and after the trimming are shown in Figure 8.4. The distribution was normalized substantially so that no further steps were necessary.

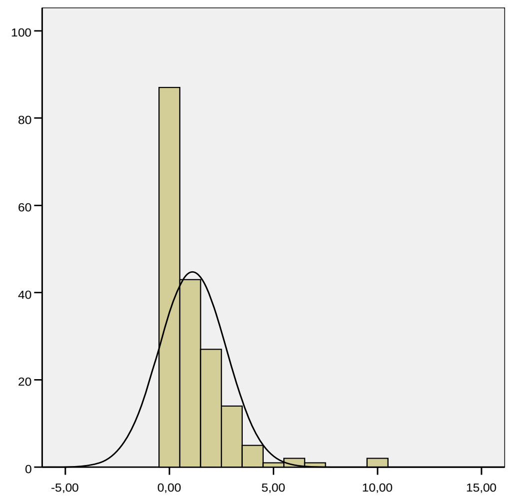


Figure 8.3

Distribution of Missings Due to Wrong Answers in SPp1 (Study 2)

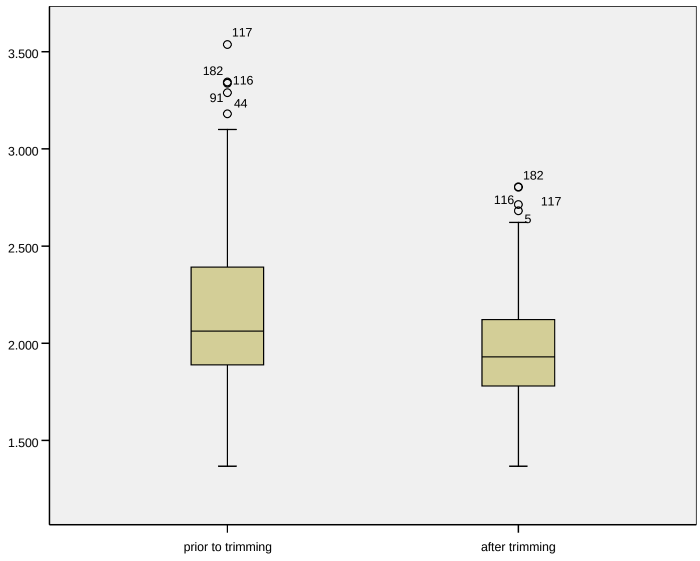


Figure 8.4

Boxplots of Reaction Time Score SPp1 Prior to and After Trimming Based on 39 Items (Study 2)

Social Perception – SPp2

This task had not been administered in Study 1. It was intended to include more complex requirements. The target stimuli thus were much more diverse and complex than in SPp1. Prior to any screening of the distributions, it was examined whether different stimuli showed substantially different reaction times. This was not the case, so the subsequent analysis relied on a compound score based on all target cues.

An inspection of the number of missings due to an absence of a response showed only very small numbers for both trials and subjects. In the next step, wrong responses were recoded into missing values. Only those items were included in the main analysis that showed a mean performance higher than the guessing rate (choice reaction time tasks with two or three choices). Six items had to be excluded because of a too large difficulty index. No reaction times were lower than 100 ms. An inspection of the distribution of a first compound score showed a slightly skewed distribution so that reaction times of single trials slower than 3 *sd* were set equal to 3 *sd*. Figure 8.5 presents the boxplots of the final scale before and after this trimming step. No further steps were necessary.

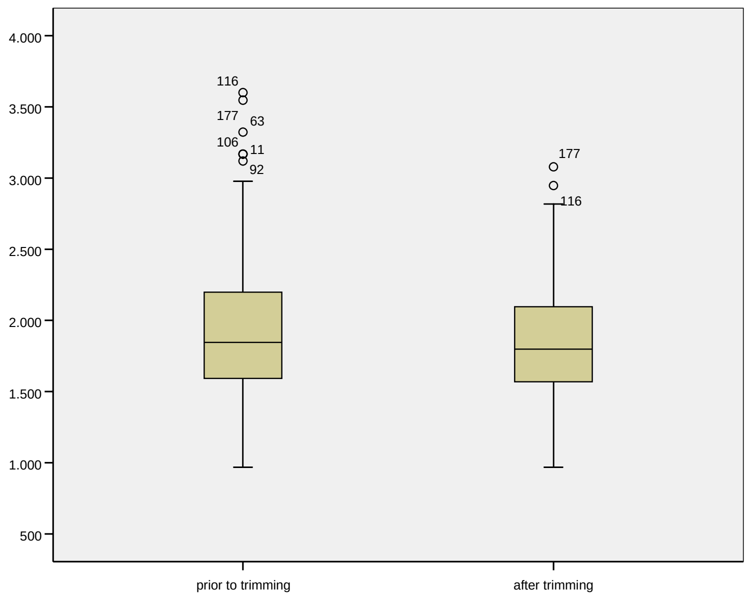


Figure 8.5

Boxplots of Reaction Time Score SPp2 Prior to and After Trimming Based on 37 Items (Study 2)

Social Perception – SPf1

The number of missing values per trial and subject were screened prior to any recoding of wrong answers or false alarms. Thus, a missing value only represented the absence of a response. Two items showed a large number of missings (81 missing data points for 182 subjects) and were excluded from further processing. However, five subjects exceeded the threshold of 15 % missings and were also omitted from the dataset. Afterwards, trials with a preceding false alarm or with a wrong response were set to missing. No wrong responses were discovered and only a very small number of false alarms occurred (i.e., between zero and three). A screening of the compound score showed no outlying values or skewed distributions. No reaction times were lower than 100 ms. Figure 8.6 presents the distribution of the final scale.

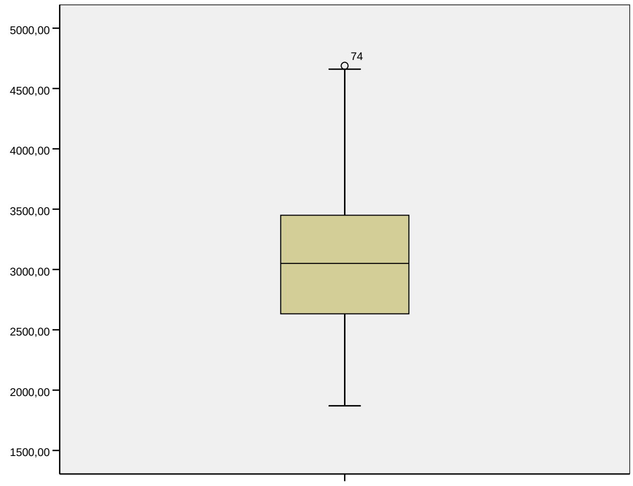


Figure 8.6
Boxplots of Reaction Time Score SPf1 Based on 28 Items (Study 2)

Social Perception – SPf2

Like the second pictorial social perception task SPp2, this task was also applied for the first time and should also depend on more complex stimulus and target material. In contrast to the video-based person perception task (SPf1), one video contained several target cues following in one row with varying distances in between. The absence of a response was attributed to a wrong or delayed perception of the relevant stimuli. Thus, the inspection of missings included the absence of a response or a wrong response. This screening step showed that five items exhibited an accuracy level below the guessing rate. These were not included in the further analysis.

The screening of the distribution of the final score showed two meaningful outliers on both sides of the distribution so that scores beyond 3 *sd* were set equal to 3 *sd* above (below) the mean. The resulting distributions are shown in Figure 8.7. No further trimming steps were conducted.

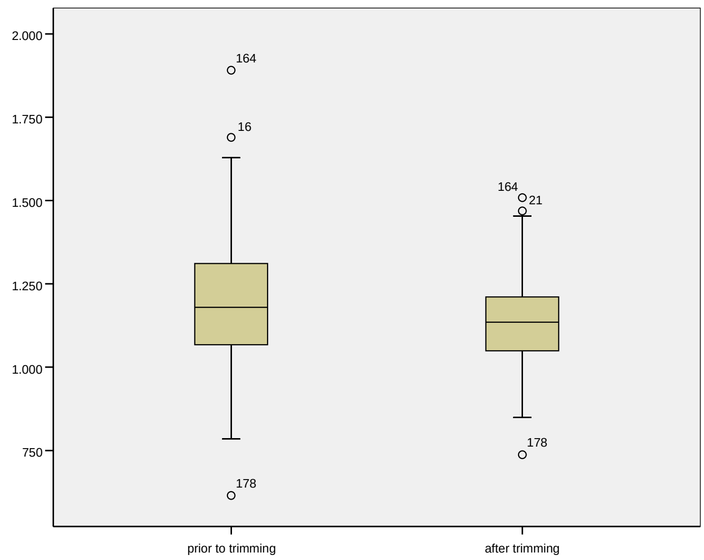


Figure 8.7

Boxplots of Reaction Time Score SPf2 Prior to and After Trimming Based on 37 Items (Study 2)

Long term memory

The task that assessed the long term memory of the information contained in the social understanding tasks was also newly constructed and inspected for missing values. One subject had to be excluded in this analysis because nearly half of the answers were missing (24 data points for 56 items).

8.4.2 Psychometric Properties and Descriptives – Research Question 1A

After the preparatory steps, the psychometric properties of the newly constructed scales were analyzed. Before building the final scales, the item-total correlations were inspected. Items were excluded that showed negative or zero item-total correlations. At the same time, it was attempted to maintain the taxonomic representativeness of the items. Most of the time, this was not problematic because many items were applied.

8.4.2.1 Social Understanding Tasks Research Question

Table 8.9 presents the results of the reliability analysis of the social understanding tasks. The scales were optimized separately for target and group consensus scoring (left and right side of the table, respectively).

Table 8.9

Reliability Analysis Social Understanding Tasks, Target and Group Consensus Scoring (Study 2)

	Target scoring				Group consensus scoring			
	Item count	r_{it} range	Cronbach's α	Range of item difficulties*	Item count	r_{it} range	Cronbach's α	Range of means
SUv**	**51**	**.070 - .458**	**.752**	**[-4.839; -.786]**	**72**	**.069 - .347**	**.765**	**[.154; .382]**
	115	-.277 - .341	.455	[-4.821; -.654]	115	-.173 - .324	.601	[.154; .473]
SUa**	**62**	**.071 - .464**	**.791**	**[-4.605; -.884]**	**103**	**.073 - .365**	**.839**	**[.149; .437]**
	152	-.313 - .344	.621	[-4.632; -.697]	152	-.173 - .350	.763	[.080; .438]
SUp**	**47**	**.058 - .400**	**.754**	**[-4.471; -.748]**	**86**	**.076 - .347**	**.792**	**[.169; .395]**
	113	-.185 - .385	.495	[-4.476; -.732]	113	-.122 - .337	.744	[.169; .403]
SUf**	**77**	**.053 - .453**	**.841**	**[-4.536; -.721]**	**101**	**.089 - .445**	**.834**	**[.160; .472]**
	124	-.289 - .417	.721	[-4.546; -.604]	124	-.107 - .420	.798	[.159; .472]
SUps**	**60**	**.068 - .462**	**.851**	**[-3.825; -1.053]**	**71**	**.120 - .391**	**.831**	**[.189; .361]**
	72	-.268 - .482	.809	[-3.824; -.974]	72	.079 - .397	.831	[.189; .361]

Note. * possible range [-6.00; .00], a higher score indicating better performance
** *upper line* of each ability domain printed in boldface indicates scales with a reduced item number after item selection, *lower line* indicating scales prior to item selection
SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

Reliability Analysis – Target and Group Consensus Scoring

In Study 1, the number of items was reduced substantially after item selection. The left columns in Table 8.9 still show a large number of items with negative item total correlations that were excluded from the scales. In the scales based on written and spoken language and on pictures, more than half of the items were excluded (56, 59, and 58 %, respectively). Only the tasks based on videos and the personality rating scales showed satisfactorily reliable scales, before item selection. All final scales were sufficiently reliable. In Study 1, the range of item difficulties had been reduced. This time, the range of item difficulties was approximately the same before and after item selection. The effect of the item selection on the taxonomy of social understanding is shown in Table 8.10. The facetted structure included an 8x4x3x2 design cross-classifying eight targets, four contents, three modalities, and two settings. On the lowest level, 192 cells resulted from the cross-classification including between one and six items. After item selection, 63 cells did not contain any more items (shaded cells in Table 8.10). This meant a reduction in representativeness of the taxonomy.

Table 8.10

Number of Items Selected for the Social Understanding Scales (Study 2)

		Targets in order of testing*																		
		1 (RF)		**2 (BS)**		**3 (CK)**		**4 (CP)**		**5 (KL)**		**6 (FB)**		**7 (HR)**		**8 (MM)**				
		pr	**pu**	**pr**	**pu**	**pr**	**pu**	**pr**	**pu**	**pr**	**pu**	**pr**	**pu**	**pr**	**pu**	**pr**	**pu**	Σ	Σ	Σ
	E	1 (2)	2 (4)	1 (3)	2 (4)	1 (5)	2 (5)	2 (3)	2 (3)	1 (4)	3 (5)	0 (1)	2 (3)	4 (4)	0 (1)	0 (4)	1 (3)	24		E_{tot}: 108
V	**C**	3 (4)	3 (4)	1 (1)	0 (1)	3 (4)	0 (3)	1 (1)	1 (1)	0 (1)	2 (3)	1 (3)	0 (4)	2 (4)	2 (2)	1 (2)	1 (2)	21	51	C_{tot}: 75
	R	1 (1)	0 (1)	1 (1)	0 (1)	1 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	1 (3)	0 (1)	2 (4)	0 (1)	6		R_{tot}: 54
	E	0 (1)	3 (5)	4 (5)	2 (3)	2 (5)	2 (5)	3 (4)	0 (1)	1 (5)	2 (5)	3 (4)	5 (6)	2 (6)	3 (4)	1 (4)	0 (1)	33		
A	**C**	3 (5)	1 (5)	1 (4)	0 (3)	2 (3)	0 (2)	2 (4)	0 (2)	0 (1)	1 (4)	0 (1)	0 (1)	2 (3)	3 (4)	0 (3)	2 (3)	17	62	
	R	0 (4)	0 (1)	0 (1)	0 (3)	0 (1)	0 (1)	2 (4)	0 (2)	2 (4)	1 (4)	3 (4)	3 (4)	0 (1)	0 (1)	1 (4)	0 (1)	12		
	E	5 (5)	0 (1)	1 (4)	1 (5)	0 (3)	1 (4)	0 (1)	1 (3)	1 (4)	0 (1)	1 (4)	1 (1)	0 (1)	4 (5)	0 (1)	3 (5)	19		
P	**C**	0 (2)	3 (3)	2 (3)	1 (3)	0 (1)	1 (1)	1 (1)	0 (1)	1 (4)	2 (4)	0 (1)	1 (2)	2 (3)	0 (3)	0 (1)	1 (3)	15	47	
	R	0 (1)	1 (1)	3 (4)	1 (4)	1 (1)	1 (1)	2 (4)	0 (1)	0 (3)	1 (1)	2 (2)	1 (1)	0 (1)	0 (1)	0 (2)	0 (1)	13		
	E	1 (1)	2 (4)	1 (1)	1 (1)	2 (3)	2 (6)	3 (4)	2 (3)	1 (1)	6 (6)	0 (1)	1 (1)	3 (4)	5 (5)	0 (2)	2 (5)	32		
F	**C**	2 (3)	1 (1)	1 (4)	2 (3)	0 (1)	0 (1)	2 (4)	1 (1)	2 (3)	2 (3)	2 (2)	0 (1)	1 (1)	4 (4)	1 (2)	1 (2)	22	77	
	R	3 (4)	1 (1)	0 (1)	1 (1)	1 (4)	4 (4)	1 (1)	1 (4)	2 (4)	1 (2)	1 (1)	1 (1)	1 (1)	1 (3)	1 (4)	3 (4)	23		
	Σ	36 (64)		27 (64)		26 (66)		27 (55)		32 (74)		28 (51)		40 (66)		21 (64)				

Note. v = written language, a = spoken language, p = pictures, f = videos, e = emotions, c = cognitions, r = relationships, pr = private, pu = public, tot = total sum
shaded cells indicating that no item is included in the final scales

No systematic effect, however, could be discovered so that no element of the facetted design had to be omitted completely. On average, 47.25% of the items were selected. Only the items assessing the relationship of the targets to other people were marginally underrepresented in the scales based on written and spoken language (i.e., only 29 and 30 % of the items were selected). Moreover, the private settings of the scenarios Katharina (KL) and Matthias (MM) were also more reduced than the remaining scales (31% and 21% of the items were selected, respective for Katharina and Matthias). Only two scenes were completely omitted by item selection, namely, the public setting of the scenario Christph (CP) based on spoken language material and the private setting of the scenario Matthias (MM) based on pictorial material.

In Study 1, the scales based on group consensus scoring were less reliable than the target scoring scales. In Study 2, less items had negative or zero item-total correlations (see

Table 8.9; i.e., between 62 and 99 %) and the scales before and after item selection were more reliable.

Descriptive Statistics – Target and Group Consensus Scoring

The descriptive statistics of the final scales are presented in Table 8.11, both for target and group consensus scoring. Remarkably, the scale means based on target scoring lay about one point lower compared to Study 1 ($m_{\text{Study 1}}$ between -1.612 and -1.05 for the content-related scales). This corresponded to the change in the rating scale from a 6-point to a 7-point scale from Study 1 to Study 2. At the same time, the standard deviation remained the same ($sd_{\text{Study 1}}$ between .364 and .486). It seemed that the difficulty of the scales based on target scoring increased from Study 1 to Study 2.

Table 8.11

Descriptive Statistics of Social Understanding Tasks Target and Consensus Scoring (Study 2)

	Task (N)	Item count	**M** all target (old / new targets)	**SD** all target (old / new targets)	Range	Skew-ness	Kur-tosis
TS	SUv (182)	51	-2.628 (-2.401 / -2.969)	.397 (.433 / .476)	[-3.77; -1.79]	-.313	-.145
	SUa (182)	61	-2.269 (-2.294 / -2.252)	.423 (.510 / .474)	[-3.66; -1.26]	-.505	.144
	SUp (182)	49	-2.252 (-2.195 / -2.315)	.422 (.458 / .548)	[-3.61; -1.27]	-.499	.632
	SUf (182)	77	-2.122 (-1.831 / -2.364)	.389 (.394 / .473)	[-3.42; -1.33]	-.506	.711
	SUps (182)	57	-2.208 (-2.181 / -2.239)	.472 (.501 / .560)	[-3.73; -1.12]	-.375	.114
GCS	SUv (182)	72	.231	.023	[.14; .29]	-.696	1.034
	SUa (182)	103	.244	.024	[.15; .29]	-.975	1.871
	SUp (182)	86	.232	.023	[.13; .28]	-1.038	2.235
	SUf (182)	101	.238	.024	[.15; .29]	-.913	1.439
	SUps (182)	71	.263	.033	[.14; .33]	-.949	1.756

Note. SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings
TS = target scoring, GCS = group consensus scoring
old targets: targets applied already in Study 1; new targets: targets only applied in Study 2

This could be attributed to two aspects. On the one hand, the sample in Study 2 was more heterogeneous in terms of age and education and showed a lower level of performance in several of the academic intelligence tasks. This might explain an increase in item difficulty in the present study. On the other hand, this effect could also be due to the test modifications and extensions after Study 1. Four more scenarios and thus, four more targets were applied in the present study which were more heterogeneous. Table 8.11 presents the scale means separately for the "old" and the "new" targets. It shows that the enhanced difficulty in Study 2 could be attributed, to some extent, to the new targets. Except for the spoken language scale, all scale means based on only the "old" targets showed a lower item difficulty. It appeared important after these findings to find a way to develop a more systematic way to investigate the item and scale difficulties. The was to be addressed again in the Discussion in Chapter 9.

Looking at the mean difficulties of the content-related scales, those based on written language seemed most difficult (-2.628) whereas the video-based scale showed the lowest difficulty (-2.122). All paired mean comparisons of the written language scale with the remaining scales were significant (t = 10.669 – 20.319; df = 181; $p < .001$). At the same time, all paired comparisons of the video-based scale with the remaining scales showed significant mean differences (t = 2.782 – 20.319; df = 181; $p < .01$). For this analysis, the alpha-level was adjusted by dividing it by the number of t-tests because they were all exploratory. This finding contradicted the results from Study 1 where the pictorial scale showed the lowest difficulty. Again, the development of a system to estimate item difficulties prior to scale construction appeared indispensible. Moreover, a replication of the present results was necessary.

The means of the scales based on group consensus scoring showed a decrease in consensus (i.e., .232 - .263 in Study 2 compared to .363 - .486 in Study 1) which, at first sight, corresponded to the increased item difficulty observed in target scoring. The effect of item difficulty on the level of consensus was to be explored in Chapter 8.4.4.3. Thus, it is not possible to make any conclusions about the interaction between the target and group consensus scoring scales at this point.

Distributions

Figure 8.8 presents the distributions of the social understanding scales based on target scoring separately for the four content-related scales. The scales were rather normally distributed as was already seen in the parameters for skewness and kurtosis in Table 8.11. The video-based scale showed some outliers which, however, did not influence the bivariate distributions because they represented different subjects. The distributions of the scales based on the personality ratings and all consensus based scales were equivalent to those displayed in Figure 8.8.

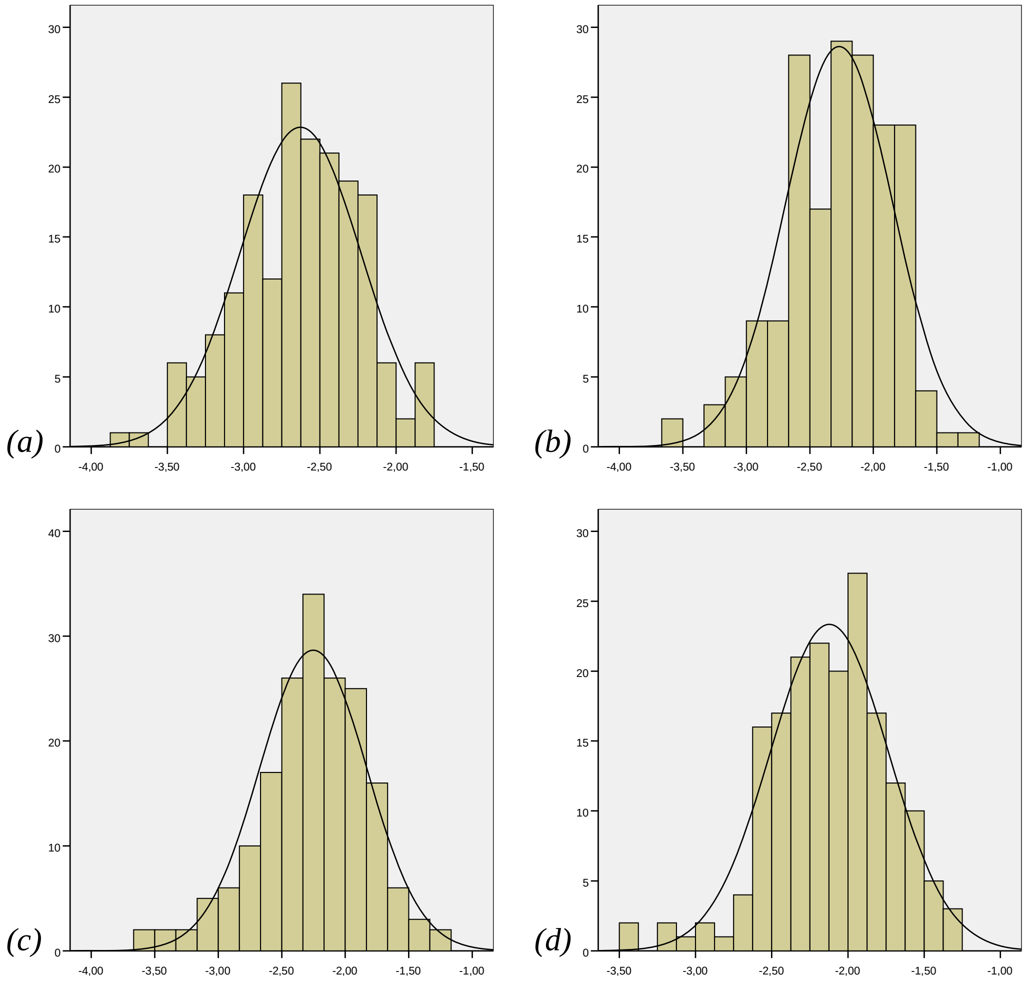


Figure 8.8

Histograms of Social Understanding Scales (Study 2)

Note. *(a)* SUv, *(b)* SUa, *(c)* SUp, *(d)* SUf

Within-Domain Correlations

As a last step, the correlations within the social understanding scales were examined. Table 8.12 presents the intercorrelations within and between the scoring methods. Within one scoring method, the scale intercorrelations were consistently large (r = .391 - .645 for target scoring and r = .512 - .709 for group consensus scoring). In contrast to Study 1, the personality ratings were also substantially correlated with the content-related scales in both scoring methods. The origin of this change in correlation size was not obvious. The only evident change between Study 1 and 2 regarded the application of 7-point rating scales also for the personality ratings (these were based on 5-point rating scales in Study 1). Whether this was the only reason for this finding could not be clarified at the moment.

The pattern of correlations between the scoring methods was not as clear as in Study 1. The target and group consensus scoring scales based on the same contents were correlated with *r* = .748 - .878 in Study 1. The remaining correlations were rather unsystematic and all were smaller than those between the same task contents. In Study 2, the correlations between the target and consensus scoring based on written language and personality ratings were substantially smaller than those based on the other contents (*r* = .314 / .466, respectively for SUv and SUps). Moreover, some correlations between different content-related scales were larger than those within one task content. The correlation pattern will be addressed again in Chapter 8.4.4.3 so that, at present, no further conclusions were undertaken.

Table 8.12

Intercorrelations of Social Understanding Scales based on Target and Consensus Scoring (Study 2)

		Target Scoring					**Group Consensus Scoring**				
		SUv	**SUa**	**SUp**	**SUf**	**SUps**	**SUv**	**SUa**	**SUp**	**SUf**	**SUps**
Target Scoring	**SUv**										
	SUa	.391**									
	SUp	.512**	.537**								
	SUf	.635**	.568**	.691**							
	SUps	.645**	.472**	.487**	.547**						
Group Consensus Sc.	**SUv**	.314**	.556**	.448**	.473**	.230**					
	SUa	.478**	.743**	.593**	.622**	.401**	.634**				
	SUp	.443**	.540**	.687**	.626**	.422**	.598**	.650**			
	SUf	.500**	.592**	.669**	.783**	.465**	.535**	.709**	.682**		
	SUps	.392**	.357**	.495**	.568**	.466**	.512**	.548**	.671**	.678**	

Note. N = 182, * α < .05, ** α < .01
SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings
shaded cells: intercorrelations between the scoring methods relying on the same content scales

8.4.2.2 Social Memory and Perception

The scales of the pictorial and video-based social memory and perception tasks were built according to the same principle as the social understanding tasks. Additionally, a reliability analysis was conducted for the long term memory task (LTM) and the PONS. The results of these analyses are presented in Table 8.13. In Study 1, the memory tasks had shown rather low reliability coefficients (i.e., .524, .136, .469, respective for SMp1, SMp2, and

SMf). In the present study, the reliabilities improved for all of the scales (.557, .455, and .649 respectively for SMp1, SMp2, and SMf). Yet, the final reliability coefficients were not sufficiently high.

Table 8.13

Descriptive Statistics of Social Memory and Perception, Long Term Memory, and PONS (Study 2)

Task (N)	Item count	M	SD	Range	Skew-ness	Kurto-sis	r_{it} range	α
SMp1 (177)	17 (18)	.698	.156	[.24; 1.00]	-.353	-.054	.087 - .327 (-.010 - .331)	.557 (.535)
SMp2 (177)	15 (20)	.497	.156	[.10; .93]	.059	-.128	.076 - .243 (-.115 - .219)	.455 (.394)
SMf (179)	29 (40)	.558	.138	[.17; .86]	-.153	-.242	.057 - .412 (-.197 - .387)	.649 (.566)
SPp1 (182)	39	1970.612	281.291	[1368.00; 2805.08]	.503	.273	.244 - .563	.906
SPp2 (182)	37	1828.090	402.535	[968.96; 3078.84]	.401	.018	.411 - .705	.954
SPf1 (177)	28	3096.509	580.436	[1869.54; 4688.30]	.472	-.250	.100 - .460	.714
SPf2 (182)	32	1134.754	124.757	[737.50; 1508.68]	.069	.402	.061 - .636	.844
LTM (181)	52 (56)	.649	.125	[.23; .91]	-.600	.527	.068 - .452 (-.045 - .438)	.816 (.804)
PONS (179)	33 (60)	.876	.071	[.64; 1.00]	-.065	.556	.068 - .351 (-.056 - .268)	.571 (.397)

Note. in parentheses are parameter before item selection based on r_{it}-inspection
SM = social memory, SP = social perception, p = pictures, f = videos; LTM = long term memory; PONS = Profile of Nonverbal Sensitivity

All four social perception tasks showed sufficient reliability coefficients (between .71 and .95), and all items correlated positively with the final score. Only the task SPf1 was at the lower end of an acceptable reliability coefficient ($\alpha = .714$). The long term memory task showed good reliability; only four items were excluded due to their item-total correlations. Interestingly, the overall performance in the long term memory task (*m* = .649) was at about the same level as the remaining social memory tasks (*m* = .497 - .757; see Table 8.14 for the means of the tasks based on written and spoken language). This was especially interesting because the subjects had not been instructed to memorize the information from the social understanding tasks. This result corresponded with the findings of Bless et al. (2004) who found better performance in memory tasks of individuals who were instructed to process the information compared to those who had only been told to memorize the information. A paired t-test showed significant mean differences for all paired comparisons of the long term memory task with the remaining social memory tasks (adjusted alpha-level of .008 for exploratory tests). The written language and the first pictorial tasks were significantly better accomplished (t = 13.376 / 3.760; df = 180 / 175; p < .001; respective for the comparisons with SMv and SMp1). All other comparisons were significant in favor of the long term memory task (t between -5.231 and -13.365; df between 176 and 180; p < .001). This overall comparison was, however, only valid to a limited extent since the general long term memory score ignored the differentiation between different task contents. This question will be addressed in more detail in Chapter 8.4.4.6.

The PONS (video version with 40 items and extended video version with 60 items) was also inspected for distribution and reliability. Nine items showed a mean performance below the guessing rate of .50 (m = .209 - .429) and were excluded. Afterwards, a reliability analysis showed that the reliability of the original video scale (40 items) was not sufficient which was, however, expected from the literature. Unfortunately, the adding of 20 items did not improve the reliability (.397 for the 60-item scale). Twenty-seven items with negative or zero item-total correlations were excluded so that the reliability of the final scale could be improved to α = .571. The distribution still showed three subjects with a performance around the guessing rate of the task (.52 - .58). Therefore, they diverged largely from the group mean (m = .876, sd = .071). The distribution parameters showed a substantial deviation from the normal distribution including these three subjects (skewness = -1.273, kurtosis = 2.940). Thus, these subjects were excluded from the analysis.

8.4.2.3 Further Measures

The psychometric properties of the remaining social intelligence tasks based on written and spoken language are presented in Table 8.14. The scale construction and preparatory data analysis of these scales are described in Seidel (2007). Except for the spoken language social memory task (SMa2), all tasks showed sufficient reliability coefficients and no meaningful deviations from the normal distribution.

Table 8.14

Descriptive Statistics of Social Intelligence Tasks Based on Written and Spoken Language (Study 2)

Task (N)	Item count	M	SD	Range	Skew-ness	Kurto-sis	r_{it} range	α
SMv1+2 (182)	48	.757	.110	[.30; .97]	-.815	1.040	.070 - .461	.844
SMa1 (182)	36	.601	.135	[.17; .88]	-.596	.281	.101 - .480	.804
SMa2 (182)	12	.498	.147	[.17; .83]	.034	-.454	-.011 - .137	.185
SPv1 (182)	60	4680.565	1346.238	[2176.12; .9503.66]	.755	.606	.351 - .782	.974
SPv2 (182)	60	2021.445	520.139	[847.75; .3689.65]	.297	.065	.535 - .774	.978
SPa1* (182)	93	-.002	.545	[-1.15; .1.63]	.534	.100	.084 - .504	.909
SPa2* (181)	46	-.010	.702	[-1.50; 2.31]	.469	.479	.265 - .609	.924

Note. * z-scores
SM = social memory, SP = social perception, v = written language, a = spoken language

Table 8.15 presents the descriptive statistics of all further measures relevant for the present research questions (i.e., the BIS scales, the speed baseline measures, the personality traits, and the self-report questionnaires). The reliability coefficients of the BIS scales were analyzed relying on the content- and operation-homogeneous parcels, respective for the operative and content abilities. The parcels for the operative ability domains consisted of equal variance components from every content domain (i.e., content-homogeneous). In turn, the operation-homogeneous parcels for the content abilities consisted of equal variance components for each operative ability. The scale descriptives were based on the aggregated scales relying on standardized-z-scores.

Table 8.15

Descriptive Statistics of Further Measures (Study 2)

Task	Item count	M	SD	Range	Skew-ness	Kur-tosis	r_{it} range	α
BIS-R* **	-	.050	.634	[-1.26; 1.67]	.367	-.509	.676 - .753	.852
BIS-M* **	-	.022	.608	[-1.68; 1.43]	-.015	-.531	.662 - .675	.817
BIS-S* **	-	.022	.579	[-1.56; 2.21]	.142	.786	.624 - .729	.823
BIS-V* **	-	.034	.556	[-1.15; 1.78]	.155	-.391	.581 - .666	.776
BIS-F* **	-	.024	.540	[-1.15; 1.44]	.218	-.498	.506 - .684	.753
BIS-N* **	-	.036	.611	[-1.21; 2.07]	.571	.257	.667 - .727	.834
SRT	50	249.984	28.426	[180.65; 356.26]	.940	1.422	.371 - .758	.964
MT	75	732.448	72.906	[543.83; 921.20]	.138	-.146	.290 - .699	.975
Readspeed	60	2656.157	552.542	[1350.85; 4083.21]	.033	-.393	.400 - .874	.977
NEO-N	12	1.653	.653	[.25; 3.58]	.523	.110	.257 - .708	.844
NEO-E	12	2.497	.562	[.67; 3.58]	-.532	.611	.206 - .697	.792
NEO-O	12	2.571	.505	[1.00; 3.67]	-.101	-.213	-.015 - .578	.690
NEO-A	12	2.611	.512	[1.00; 3.92]	-.404	.373	.173 - .586	.761
NEO-C	12	2.675	.636	[1.08; 3.92]	-.330	-.428	.390 - .678	.867
SB Questionnaire	40	2.722	.336	[1.75; 3.54]	-.217	-.116	.168 - .575	.876
SEIS	33	3.780	.378	[2.67; 4.88]	-.215	.013	.177 - .559	.848
Altruism	12	.637	.186	[.08; 1.00]	-.344	-.236	.146 - .354	.591
Empathy – Compassion	7	3.837	.622	[2.14; 5.00]	-.258	-.288	.300 - .565	.735
Empathy – Perspective taking	7	3.479	.675	[1.86; 4.86]	-.299	-.397	.320 - .638	.753
Depression	24	1.832	.429	[1.04; 3.08]	.614	-.229	.169 - .631	.882

Note. N = 182, * z-scores
** reliability analysis of the BIS scales based on content- and operation-homogenous parcels, four parcels for BIS-R, three parcels for BIS-S, -M, -V, -F, -N
R = Reasoning, M = Memory, S = Speed, V = verbal abilities, F = figural-spatial abilities, N = numerical abilities, NEO-N = Neuroticism, NEO-E = Extraversion, NEO-O = Openness, NEO-A = Agreeableness, NEO-C = Conscientiousness, SB = social behavior, SEIS = Schutte Emotional Intelligence Scale

The reliability coefficients were good across all measures. Only one self-report questionnaire showed no satisfactory reliability coefficient (i.e., Altruism, $\alpha = .591$). This scale, however, was only based on 12 items. The two subscales of the Empathy questionnaire (i.e., empathic compassion and cognitive perspective taking) were analyzed separately because the subscales were only marginally intercorrelated ($r = .130$; N = 182; n.s.).

8.4.3 Construct Validity

In the context of examining the construct validity of social intelligence, the internal structure of social intelligence should be investigated (research question 2A) as a first step. As a second step, the relationship of the social intelligence tasks to the PONS should be explored in order to investigate the convergent construct validity (research question 2B). As a third step, social intelligence should prove divergent construct validity with academic intelligence and personality (research question 2C).

8.4.3.1 Structure of Social Intelligence – Research Question 2A

The structure of social intelligence should be investigated by relying on correlational (research question 2A1) and confirmatory factor analysis (research question 2A2). As a last step, the faceted structure of social intelligence should be explored (research question 2A3).

Correlational Results – Research Question 2A1

Table 8.16 presents the correlations for the social intelligence tasks. The correlations within the social understanding scales are not included in this table since they were already presented in Table 8.12. As Table 8.12 shows, the scales yielded substantial within-domain correlations (*r* between .391 and .691). Moreover, Table 8.16 shows only marginal and unsystematic correlations between the social understanding tasks and social memory and perception. The social memory tasks were rather consistently intercorrelated (r = .137 - .630) with the smaller and nonsignificant correlations of the spoken language task SMa2. This task, however, had a poor reliability coefficient of α = .185 which could be responsible for the small correlations. The correlations between social memory and social perception were heterogeneous ranging from r = -.471 to r = .073; negative correlations represented a hypothesis-conforming relationship because of the negatively coded reaction time scores of social perception. The correlations within the social perception tasks were also heterogeneous but all were positive in sign. In the present study, four tasks were added (i.e., SPv2, SPa2, SPp2, and SPf2). In comparison to Study 1, most of the tasks were meaningfully and significantly correlated across the different content domains.

Table 8.16 *Correlations Between Social Intelligence Tasks (Study 2)*

	SUv	SUa	SUp	SUf	SMv	SMa1	SMa2	SMp1	SMp2	SMf	SPv1	SPv2	SPa1	SPa2	SPp1	SPp2	SPf1	SPf2
SMv	.098	.126	.165*	.155*							-.291**	-.191**	.043	.003	-.024	.030	.013	.013
	(182)	(182)	(182)	(182)														
SMa1	.086	.206**	.152*	.127	.630**						-.140	-.163*	.125	-.131	.053	.052	.062	.051
	(182)	(182)	(182)	(182)	(182)													
SMa2	.000	.048	.003	-.068	.170*	.203**					-.173*	-.141	.010	-.043	-.091	-.079	-.192*	.007
	(182)	(182)	(182)	(182)	(182)	(182)												
SMp1	.029	.074	.163*	.109	.327**	.284**	.122				-.161*	-.063	-.173*	.000	-.207**	-.079	-.206**	-.180*
	(177)	(177)	(177)	(177)	(177)	(177)	(177)											
SMp2	.075	.037	.154*	.109	.377**	.341**	.152*	.249**			-.074	-.052	-.041	-.008	-.097	.002	-.116	-.064
	(177)	(177)	(177)	(177)	(177)	(177)	(177)	(173)										
SMf	.016	.054	.068	-.026	.406**	.454**	.137	.189*	.326**		-.118	-.101	.099	-.167*	-.186*	-.010	-.141	-.085
	(179)	(179)	(179)	(179)	(179)	(179)	(179)	(174)	(176)									
SPv1	-.062	-.056	-.083	-.123	-.471**	-.283**	-.201**	-.225**	-.147	-.103		.333**	.098	.143	.081	.167*	-.008	.006
	(182)	(182)	(182)	(182)	(182)	(182)	(182)	(177)	(177)	(179)								
SPv2	-.125	-.081	-.122	-.171*	-.425**	-.326**	-.178*	-.174*	-.148*	-.077	.660**		.019	.183*	.053	.022	-.142	-.002
	(182)	(182)	(182)	(182)	(182)	(182)	(182)	(177)	(177)	(179)	(182)							
SPa1	-.002	-.132	-.086	-.047	-.040	.073	.016	-.216**	-.076	.066	.113	.217**		.150*	.151*	.066	.158*	.406**
	(182)	(182)	(182)	(182)	(182)	(182)	(182)	(177)	(177)	(179)	(182)	(182)						
SPa2	.051	-.045	-.055	.007	-.109	-.190*	-.053	-.065	-.055	-.169*	.317**	.361**	.292**		-.038	.167*	.043	.178*
	(181)	(181)	(181)	(181)	(181)	(181)	(181)	(177)	(176)	(178)	(181)	(181)	(181)					
SPp1	.025	-.040	-.039	-.044	-.067	.006	-.074	-.257**	-.127	-.198**	.206**	.217**	.318**	.141		.273**	.374**	.096
	(182)	(182)	(182)	(182)	(182)	(182)	(182)	(177)	(177)	(179)	(182)	(182)	(182)	(181)				
SPp2	-.048	-.059	-.118	-.059	-.032	.008	-.083	-.125	-.030	-.028	.256**	.174*	.176*	.256**	.366**		.210**	.236**
	(182)	(182)	(182)	(182)	(182)	(182)	(182)	(177)	(177)	(179)	(182)	(182)	(182)	(181)	(182)			
SPf1	.119	-.078	-.033	.095	-.067	.006	-.197**	-.236**	-.144	-.146	.137	.090	.235**	.136	.397**	.258**		.191*
	(177)	(177)	(177)	(177)	(177)	(177)	(177)	(172)	(172)	(174)	(177)	(177)	(177)	(176)	(177)	(177)		
SPf2	.071	-.008	-.068	.039	-.038	.023	.009	-.210**	-.086	-.094	.116	.128	.466**	.257**	.195**	.289**	.237**	
	(182)	(182)	(182)	(182)	(182)	(182)	(182)	(177)	(177)	(179)	(182)	(182)	(182)	(181)	(182)	(182)	(177)	

Note. pairwise N in parentheses, * p < .05, ** p < .01, correlations of baseline-controlled measures above diagonal, SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

At first sight, the task modifications and extensions conducted after Study 1 seemed to have succeeded better in assessing a coherent ability domain of social perception. Originally, the intention was to control for the speed baselines of the social perception tasks in order to exclude this variance. Three baseline speed measures were applied (i.e., simple reaction time, a mouse speed, and a reading speed task). For example, the written language social perception tasks were supposed to rely on the reading speed and on the simple reaction time of participants. Other tasks were supposed to rely on the mouse speed and the simple reaction time (i.e., SPp1); all other tasks should only rely on the simple reaction time. Thus, the relevant baseline measures differed between the tasks. The baseline measures were intercorrelated to a medium extent ($r_{\text{SRT, MT}} = .313$; $r_{\text{SRT,Readspeed}} = .309$; $r_{\text{MT, Readspeed}} = .181$). The zero-order correlations between the social perception tasks and the baseline measures showed an equivocal pattern (see Appendix G for the correlations of the social perception tasks with the baseline measures and within the baseline measures). All tasks were correlated with SRT (r = .198 - .433), most of the tasks with MT (r = .088 - .491) and the reading speed (r = .139 - .792).

In order to control for the baseline, all three baseline measures were entered into a multiple regression analysis to predict each social perception task, and the standardized residuals were saved. Otherwise, different variance parts would have been removed. The intercorrelations of the standardized residuals are presented above the diagonal in Table 8.16. Some of the previously substantial correlations were reduced to nearly zero (e.g., SPv2 with SPa1 or SPp1), some maintained to a substantial extent (e.g., SPp1 with SPf1; SPa1 with SPf2). The social understanding task correlations were zero prior to and after the baseline correction (see Appendix G for the correlations which were not included in Table 8.16 due to limited space). The correlations with the social memory domain are also presented above the diagonal in Table 8.16. Most of them decreased to a substantial extent after controlling for the speed baseline measures, particularly those between the written language social perception tasks and the social memory tasks.

In summary, it was clear that the speed baseline measures represented a substantial amount of variance within the social perception measures. The amount of explained variance by all three baseline measures ranged from $R^2 = .060$ for SPf1 to $R^2 = .638$ for the task SPv2. It was a very conservative strategy to control for the total baseline variance in all of the social perception tasks. At the same time, it seemed clear that this controlled variance was not of interest in terms of assessing the *social* perceptual components in the perception tasks.

Therefore, the subsequent analysis relied on the standardized residuals of the social perception tasks.

Confirmatory Factor Analysis – Research Question 2A2

Confirmatory factor analysis should investigate the structure of social intelligence as postulated in the performance model of Weis and Süß (2005). Prior to the factor analysis, measurement models for the single operative ability domains were established. These supported social understanding and social memory as coherent ability factors with positive and meaningful loadings of all indicators. A measurement model for the social perception factor could not be established involving all indicators. The written language tasks did not load on the social perception factor, and all other indicators showed positive loadings (between .22 and .49). Therefore, a model involving the remaining variables was postulated. Nevertheless, the subsequent confirmatory factor analysis was conducted by partially including the social perception tasks except for those based on written language.

The rationale of the analysis postulated several general, structural, and hierarchical models of social intelligence. Table 8.17 presents the rational and the model fit statistics. Two general factor models were postulated with Model A involving the indicators of all three ability domains and Model B only relying on the social understanding and memory tasks. Both models showed poor data fit with a slight advantage of the model excluding the social perception tasks from the analysis (CFI = .461; $\chi^2 = 280.180$; $p < .001$). In Model A, the social perception tasks did not load on the general factor (-.09 - .02). The loadings of the social understanding and memory tasks on the general factor in Model B were all significant and positive in sign (.25 - .76).

In the next step, four structural models were established postulating three (two) operative ability factors (social understanding and social memory with or without social perception). The ability factors were either correlated or not (Model C and E respective for three and two correlated factors and Model D and F respective for three and two uncorrelated factors). Data fit of all structural models was substantially better than that of the general factor models. Both of the two-factor models of social understanding and memory (Model E and F), however, showed a substantially better data fit than the three-factor models (Model C and D) indicating that social perception as assessed in the present study did not fit into a structural model of social intelligence. Thus, all further models were established without the social perception tasks.

Table 8.17

Fit Statistics for Confirmatory Factor Analyses of the Structure of Social Intelligence (Study 2)

Model	χ^2	DF	p (χ^2)	CFI	RMSEA	SRMR	CI RMSEA***
A: General factor model (with SU, SM, and SP variables)*	451.026	104	<.001	.371	.142	.138	[.128; .155]
B: General factor model (with SU and SM variables)**	280.180	35	<.001	.461	.202	.167	[.180; .224]
C: 3-Factor structural model (SU, SM, and SP correlated)*	159.158	101	<.001	.895	.059	.074	[.041; .076]
D: 3-Factor structural model (SU, SM, and SP uncorrelated)*	165.906	104	<.001	.888	.060	.088	[.042; .076]
E: 2-Factor structural model (SU and SM correlated)**	31.072	34	.612	1.000	.000	.043	[.000; .049]
F: 2-Factor structural model (SU and SM uncorrelated)**	36.075	35	.418	.998	.013	.075	[.000; .057]
G: Hierarchical model[a] (Schmid-Leiman)**	22.469	26	.663	1.000	.000	.046	[.000; .050]
H: 2 content factors (language and nonlanguage-based)**	279.340	34	<.001	.460	.205	.166	[.183; .227]
I: Facetted model: 2 operative and 2 content factors (both correlated)**	14.396	23	.915	1.000	.000	.030	[.000; .024]

Note. * N = 167, ** N = 172, *** CI = 90%, [a] two error terms constrained to be equal
SU = social understanding, SM = social memory, SP = social perception
all social perception factors composed without written language tasks

The structural models relying on two operative ability factors showed very good data fit; the models were nested. A χ^2-differences test showed a significantly better fit of Model E with two correlated ability factors (χ^2-difference = 5.003, df = 1, p < .05). Figure 8.9 presents the standardized solution of Model E. Social understanding and memory were significantly correlated with r = .20 whereas this correlation was substantially lower than that in Study 1 ($r_{SU, SM}$ = .35). All indicators loaded positively and significantly on the respective ability factors. Thus, the structure of social intelligence as supported in Study 1 could be replicated in the present study.

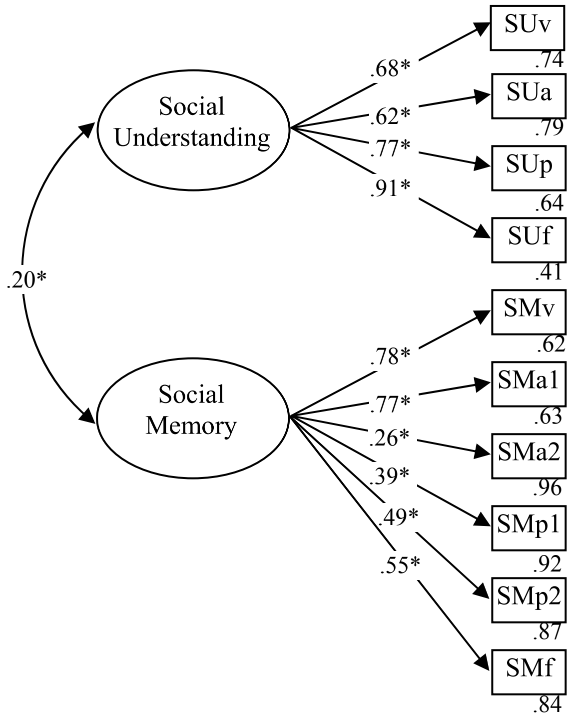


Figure 8.9

Standardized Solution of the Structural Model of Social Intelligence (Model E) with Two Correlated Operative Ability Domains (Study 2)

Note. CFI = 1.000; χ^2 = 31.072 (p=.612); * p < .05; error terms are enclosed with the manifest variables
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos

The next model (Model G) postulated a hierarchical model of social intelligence with two uncorrelated operative ability factors and one general social intelligence factor (Schmid-Leiman solution). The error term of the video-based social memory task was found to be at the lower bound and was constrained to be equal with the error term of the pictorial social memory task SMp2 (see Figure 8.10 for the standardized solution). The loading pattern was rather consistent for the operative ability factors with only positive and significant loadings on social understanding and almost only significant loadings on social memory. Only the second spoken language social memory task (SMa2) did not load significantly on the social memory factor which could have been due to the low task reliability. The loadings on the social memory factor were all negative in sign which was, however, not problematic because all factors were postulated to be orthogonal. The loadings on the general social intelligence factor were rather heterogeneous but all pointed to the same direction which was not found in Study 1. The general model fit was very good although (CFI = 1.000; χ^2 = 22.469, p=.663) it was seen as problematic to introduce an equality constraint. Thus, Model G pointed towards the hierarchical character of social intelligence although the model had to be interpreted with care due to the measurement problems. Moreover, the loadings on the general factor were partially small. This was, however, no surprise since the factor intercorrelation between social understanding and memory was rather small. Models E and G were nested. The χ^2-differences test showed a nonsignificant difference (χ^2-difference = 8.603, df = 8, n.s.) so that Model E is the preferred model due to the principle of parsimony.

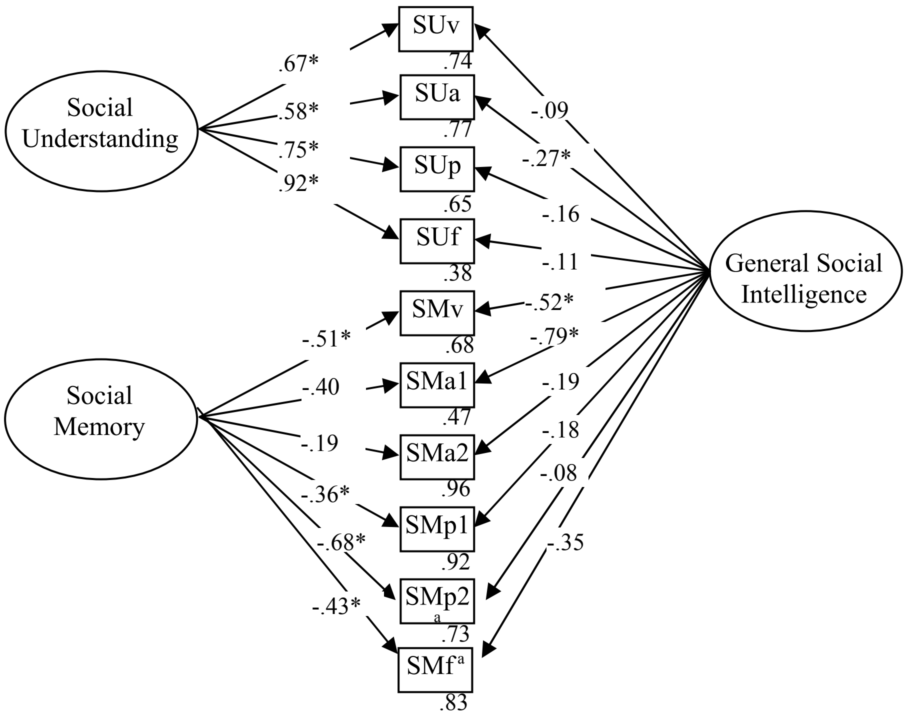


Figure 8.10

Standardized Solution of the Hierarchical Model of Social Intelligence (Model G, Schmid-Leiman) (Study 2)

Note. CFI = 1.000; χ^2 = 22.469, p=.663; * p < .05; error terms are enclosed with the manifest variables
[a] error terms constrained to be equal
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos

Exploring the Content Factors - Research Question 2A3

As a last step, the faceted structure of social intelligence should be investigated. Therefore, first a model relying only on the content-related factors should be established. The content model (Model H) only differentiated between a language-based content factor and a content factor based on language-free material. The tasks based on written and spoken language should load on the language-based factor, and the pictorial and video-based tasks should load on the language-free content factor. It was not distinguished further into the four different task contents or any other differentiation because the written language and the video-based content factors would then only rely on two indicators. The model showed poor fit statistics (CFI = .460; χ^2 = 279.340, p < .001; see Table 8.17) and the loading pattern was rather heterogeneous.

This fact notwithstanding, Model I postulated a faceted model of social intelligence according to the faceted design of the MTMM matrix underlying the SIM (see Table 8.17 for the fit statistics). It postulated two correlated operative factors (i.e., social understanding and memory as established in the structural models) and two correlated content factors (i.e., language-based and language-free). No measurement problems were encountered during optimization. The model showed very good fit statistics (CFI = 1.000; χ^2 = 14.396, p=.915). Figure 8.11 shows the standardized solution of Model I. The factor intercorrelations and loadings of the operative ability factors were similar to those in the two-factor structural model of social intelligence ($r_{SU, SM}$ (Model E) = .20; $r_{SU, SM}$ (Model I) = .25).

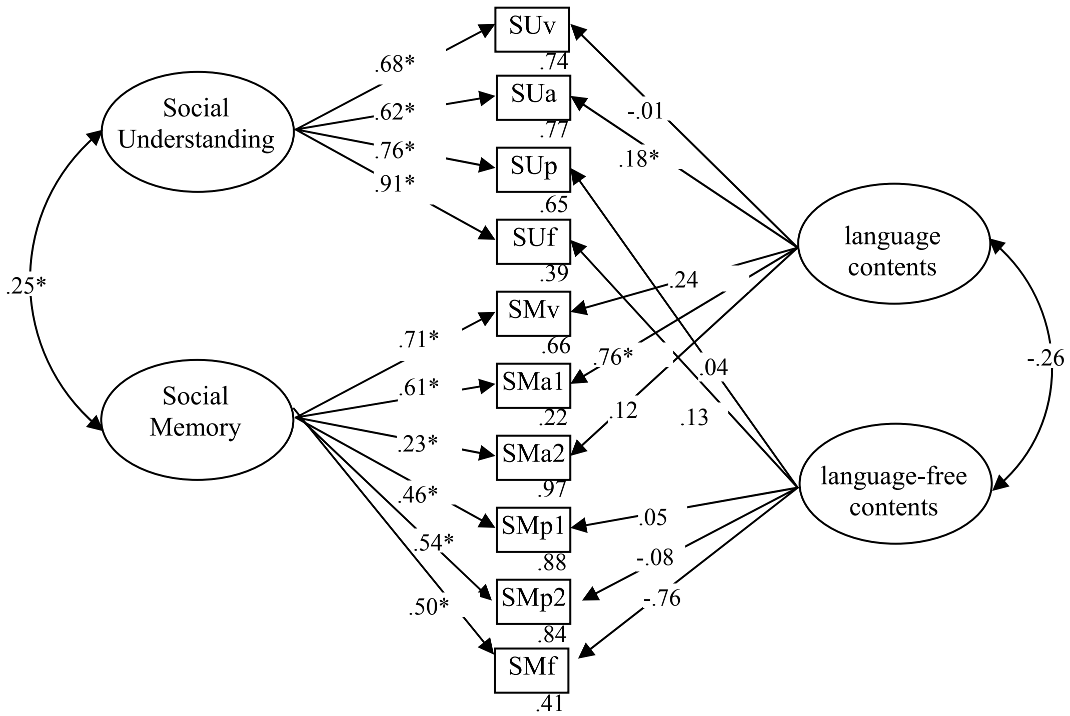


Figure 8.11

Standardized Solution of the Faceted Model I (Model I) of Social Intelligence (Study 2)

Note. CFI = 1.000; χ^2 = 22.469, p=.663; * p < .05; error terms are enclosed with the manifest variables
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos

All loadings on the operative factors reached significance and were positive in sign. The loadings on the content-related factors were very heterogeneous so that the factors were hardly interpretable (-.01 - .76 for language-based contents and -.13 - .76 for language-free contents). The two factors were negatively intercorrelated (*r* = -.26) which was attributed to the negative loadings on the language-free content factor. Models E, G, and I were nested (i.e., the structural model, the hierarchical Schmid-Leiman solution and the faceted model). The difference between Model G and I was significant (χ^2-difference = 8.073, df = 3, $p < .05$) with an advantage of the faceted model (Model I). Moreover, no measurement problems had occurred for Model I so that the hierarchical nature of social intelligence had to be questioned. However, comparing Model I with the structural social intelligence model (Model E), the difference turned out to be not significant (χ^2-difference = 16.676, df = 11, n.s.). Thus, the more parsimonious Model E was preferred.

The results from confirmatory factor analysis basically replicated the results from Study 1. Thus, the postulated structure of social intelligence could be supported a second time relying on a totally different sample. The structural model showed the best fit statistics and at the same time, was the most parsimonious model. Factor loadings were equal to those in Study 1 whereas the factor intercorrelation was substantially smaller. The hierarchical model was also replicated with still, however, some measurement problems. Moreover, data supported a faceted model of social intelligence postulating two correlated operative ability factors and two content factors with heterogeneous loadings of the respective task contents. The content factors were hardly interpretable because of the loading pattern. There did not seem to be meaningful common content-related variance in the tasks; this factor rather seemed to allocate “residual” variance.

8.4.3.2 Convergent Construct Validity – Research Question 2B

The Profile of Nonverbal Sensitivity (Rosenthal et al., 1979) was utilized in the present study. The PONS assessed nonverbal sensitivity which was supposed to be related to social understanding as defined in the present work. Only the video-based items were selected for testing because the test was in the English language. The correlations with the social understanding tasks ranged between *r* = -.004 and *r* = -.163 (N = 179; $p < .05$). The negative sign did not conform to the coding so that no evidence for the convergent construct validity with the PONS was provided. The social memory tasks, however, correlated with the PONS between $r = .042$ and $r = .181$ (N = 179; $p < .05$), the social perception tasks with r between -.119 and .101. These correlations were in the expected direction but still rather low.

In turn, empirical evidence of existing studies could not prove the convergent validity of the PONS itself (Bernieri, 2001; Buck, 1983). Only Bänziger (2005) found a substantial correlation between the PONS and the MERT. Moreover, the psychometric problems surrounding the PONS had to be accounted for as the test reliability was rather low with $\alpha =$ .57. Moreover, the distribution of the PONS was skewed due to a ceiling effect ($m = .876$; $sd = .71$ for a maximum score of 1). An inspection of possible influential cases, however, did not show any substantial influences by single values.

8.4.3.3 Divergent Construct Validity – Research Question 2C

Relationship to Academic Intelligence (Cell Level) – Research Question 2C1

The rationale of the analysis to investigate the divergent construct validity with academic intelligence was similar to that of Study 1. Social intelligence should be discriminable from academic intelligence as assessed by the BIS-Test (Jäger et al., 1997). The BIS-Test provided the opportunity to compare the two constructs on several hierarchical levels (i.e., the cells, the broad ability factors, a general level, and in terms of the faceted structure).

As a first step, the correlations between the social intelligence tasks and the BIS cells were analyzed. The cells cross-classified the operations BIS-Reasoning, -Memory, and –Speed, and the contents BIS-verbal, -figural-, and –numerical abilities. Correspondingly, the cells were labeled respective to the first letter of the operative and the content ability factors (e.g., the BIS-Reasoning cell based on figural contents was labeled cell RF). The BIS cells and social intelligence tasks were supposed to be on the same hierarchical level and both represented cross-classifications of operative ability and content domains. Table 8.18 presents the intercorrelations. The social understanding tasks were only marginally correlated with the BIS cells and only showed significant correlations with verbal reasoning ($r = .121$ - $.233$) and numerical speed ($r = .054$ - $.229$). The social memory tasks were systematically and most highly correlated with all BIS-Memory cells. The tasks SMv was also correlated with the cells RV, RN, SV, and SN. Only the second spoken language task (SMa2) was not meaningfully correlated with any BIS cell which could be attributed to the tasks' low alpha coefficient of .182. In Study 1, the overlap with the BIS-Reasoning and –Speed domains was substantially larger for all social memory tasks. This was attributed to a high task difficulty and too much restricted presentation and answering times of the social memory tasks. Thus, at first sight, the task modifications in Study 2 seemed to have succeeded in reducing variance related to speed requirements and the use of written language.

Table 8.18

Correlations of Social Intelligence Tasks with BIS Cells (Study 2)

	BIS cells								
	RV	**RF**	**RN**	**MV**	**MF**	**MN**	**SV**	**SF**	**SN**
SUv (182)	.222**	.109	.154*	-.038	-.030	-.029	.127	.045	.202**
SUa (182)	.233**	.166*	.121	.144	.084	.100	.020	.115	.170*
SUp (182)	.121	-.009	-.071	.084	-.007	.018	.121	.027	.054
SUf (182)	.208**	.084	.076	.073	.018	.033	.117	.135	.229**
SMv (182)	.449**	.156*	.305**	.498**	.340**	.349**	.332**	.132	.292**
SMa1 (182)	.349**	.082	.120	.524**	.253**	.302**	.288**	.112	.194**
SMa2 (182)	.032	.057	.037	-.025	.105	.134	.048	.027	-.007
SMp1 (177)	.181*	.105	.082	.314**	.220**	.319**	.066	.091	-.013
SMp2 (177)	.189*	.098	.110	.260**	.244**	.234**	.152*	.112	.123
SMf (179)	.045	.044	.035	.297**	.259**	.247**	.074	.083	-.068
SPv1 (182)	-.153*	-.068	.006	-.132	-.112	-.118	-.079	-.025	.030
SPv2 (182)	-.131	-.041	.115	-.165*	-.012	-.112	-.090	-.084	.078
SPa1 (182)	.045	.008	.144	.069	-.057	.009	-.011	-.052	.103
SPa2 (181)	-.071	-.091	.026	-.047	-.093	.061	-.023	-.092	.115
SPp1 (182)	.021	-.008	.100	-.036	-.194**	-.224**	-.059	-.291**	.034
SPp2 (182)	.047	-.112	-.057	.042	-.105	.014	.062	-.086	.039
SPf1 (177)	.135	.058	.115	.038	-.194**	-.045	.117	-.145	.189*
SPf2 (182)	.137	.029	-.013	-.072	-.063	-.037	-.012	.015	.072

Note. pairwise N in parentheses in left column, * p < .05, ** p < .01
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos; R = BIS-Reasoning, M = BIS-Memory, S = BIS-Speed, V = verbal, F = figural-spatial, N = numerical

The social perception variables in the present analysis were represented by the residuals after the baseline speed variance was partialled out. They showed only some unsystematic correlations across the BIS cells. The two person perception tasks based on pictures and videos were correlated with the figural memory cell; the pictorial task was also related to figural speed as would be expected. The correlations with the BIS cells of the original tasks were substantially larger particularly for the written language tasks (see Appendix G for the complete correlation matrix). SPv1 correlated with the BIS cells with r = (-.430) – (-.175) prior to the baseline control, and SPv2 with r = (-.475) – (-.190). The highest correlations for both tasks were discovered within the verbal reasoning cell. The correlations disappeared completely after partialling out the baseline variance. The same change of correlations was discovered for the spoken language tasks. Only the correlations of the pictorial and video-based social perception tasks of person perception (SPp1 and SPf1) were maintained to a meaningful extent.

In summary, the correlative pattern supported the independency of the social understanding and the social perception tasks from the BIS cells. It also pointed to successful task modifications of the social memory tasks. The overlap between these tasks and the BIS-Memory cells, however, was still rather large so that it was not clear whether social memory was discriminable from BIS-Memory.

Relationship to Academic Intelligence (Broad Ability Factor Level) – Research Question 2C1

The BIS cells were partly substantially intercorrelated (see Appendix G; between $r = .092$ for SF with RN and $r = .686$ for SN with RN). Therefore, the construct overlap on the level of the broad ability factors should be investigated by the use of confirmatory factor analysis. Before any analysis including social and academic intelligence ability factors were conducted, the structure of the operative and the content ability facet of the BIS should be empirically supported by confirmatory factor analysis. Therefore, two models were established, one postulating three correlated operative ability factors (Model J; BIS-Reasoning, -Memory, and –Speed), and the other three content ability factors (Model K; BIS-Numerical, -Verbal, and –Figural-Spatial abilities). Table 8.19 presents the model fit statistics. The models relied on homogeneous parcels including three per factor. The model based on the operative factors showed reasonable data fit (CFI = .980; $\chi^2 = 38.544$, $p < .05$). All parcels loaded significantly and positively on the respective factors (.71 - .87), and the factors were significantly correlated with *r* between .47 and .58. The content-factor model did not fit the data quite as well (CFI = .942; $\chi^2 = 64.623$, $p < .01$) although all factors showed coherent loadings (.64 - .80) and were significantly correlated with $r = .67 - .75$.

Table 8.19

Fit Statistics for Confirmatory Factor Analyses of the Divergent Construct Validity of Social with Academic Intelligence (Study 2)

Model	χ^2	DF	p (χ^2)	CFI	RMSEA	SRMR	CI RMSEA*
J: BIS-Operative factors[a]	38.544	24	<.05	.980	.058	.040	[.018; .090]
K: BIS-Content factors[a]	64.623	24	<.01	.942	.097	.058	[.068; .125]
L: SI-BIS 4 correlated ability factors[b]	144.078	98	<.01	.954	.052	.059	[.033 .070]
M: SI-BIS, SU uncorrelated with BIS[b]	146.884	100	<.01	.954	.052	.066	[.033; .070]
N: SI-BIS, SU uncorrelated with BIS[b] and SM	152.625	101	<.01	.949	.055	.077	[.036; .071]
O: BIS facetted structure based on cells[a]**	29.792	17	<.05	.977	.043	.064	[.021; .102]
P: SI-BIS facetted structure; SI integrated in BIS[b]***	20.137	12	.065	.979	.061	.043	[.000; .106]
Q: SI-BIS facetted structure; SI as additional content-factor[b]***	11.452	10	.323	.996	.028	.024	[.000; .088]
R: SI-BIS facetted structure; SI as additional operative factor[b]***	23.253	10	<.01	.966	.086	.038	[.040; .131]
S: 2-Factor structural model (residuals of SU and SM, correlated)[b]	32.241	34	.554	1.00	.000	.044	[.000; .051]
T: SI facetted model with SU and SM correlated and language and not language-based factors correlated (based on residuals)[b]***	22.036	25	.634	1.00	.000	.034	[.000; .052]

Note. [a] N = 182, [b] N = 172, * CI = 90%, ** four equality constraints, *** two equality constraints, SI = social intelligence, SU = social understanding, SM = social memory

Turning to the investigation of construct overlap, the rationale of confirmatory factor analysis first established three models which examined whether the operative ability factors of social and academic intelligence were separable or not. All three models were based on the BIS operative structure (i.e., only BIS-Reasoning and BIS-Memory) and the structural model of social intelligence with correlated social understanding and memory factors (Model E in Table 8.16). Model L introduced correlations between all four ability factors (i.e., BIS-Reasoning and –Memory; social understanding and memory). The model showed reasonable data fit and positive loadings of all of the indicators on the respective factors. The factor social understanding, however, was only marginally and not significantly correlated with the

BIS-Reasoning and –Memory factors (r = .14 and .02, respectively). Therefore, the next model (Model M) omitted the factor intercorrelations between the BIS and social understanding. The model fit did not change substantially (CFI = .954; $\chi^2 = 146.884$, p < .01) and the χ^2 difference test was not significant (χ^2-difference = 2.806, *df* = 2, n.s.) so that the more parsimonious model M was accepted as the better solution. The model is presented in Figure 8.12. All indicators loaded substantially on their respective ability factors. Social memory was meaningfully correlated with BIS-Memory (*r* = .67) and also with BIS-Reasoning (*r* = .39). Compared to Study 1, the overlap of social memory with BIS-Reasoning was reduced as was intended by the task modifications (*r* = .46 in Study 1).The correlation, however was still of substantial size.

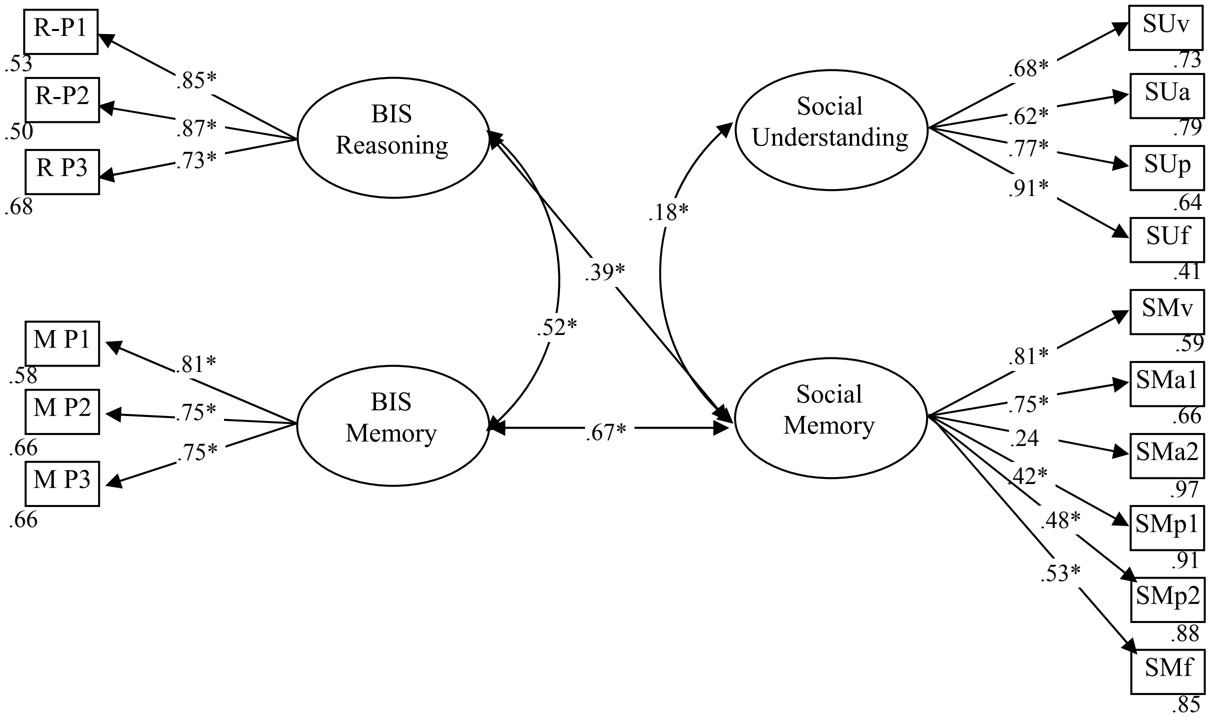


Figure 8.12

Standardized Solution of Model M Showing Factor Intercorrelations between Social and Academic Intelligence (Study 2)

Note. CFI = .954; $\chi^2 = 146.884$, p < .01; * p < .05; error terms are enclosed with the manifest variables
R = BIS-Reasoning, M = BIS-Memory, P1-3 = parcel 1-3; SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

Social understanding and memory were only marginally intercorrelated (*r* = .18) as was the case in the structural model of social intelligence (*r* = .20 in Model E in Figure 8.9). It was not clear whether the two social ability factors represented two ability factors of the same ability construct. Therefore, a third model omitted the correlations within the social intelligence factors (Model N). The model fit worsened a little bit and the χ^2 -difference test showed a significantly better data fit for Model M with correlated social intelligence factors (χ^2-difference = 5.741, df = 1, $p < .05$).

Relationship to Academic Intelligence (Integration of Social Intelligence into the Faceted BIS Structure) – Research Question 2C3

The next series of confirmatory factor analyses investigated whether social intelligence fit into the faceted structure of the BIS-Model. The faceted structure of the BIS-Model could only be established by examining the cells of the BIS that cross-classify three operative (BIS-Reasoning, -Memory, and -Speed) and three content ability factors (BIS-Verbal, -Numerical, and -Figural-spatial abilities). The factors were supposed to be intercorrelated. The forth operative ability factor of creativity was not included in the design of the present study. Moreover, the cells of the BIS-Reasoning domain did not equate exactly with the original test conceptualization since only three out of five tasks per cell were applied in the present study. This restricted faceted model of the BIS (Model O in Table 8.19) showed a reasonable fit to the data (CFI = .977; χ^2 = 29.792, $p < .05$) but needed the introduction of four equality constraints because of several error terms at a lower bound.

However, the BIS-Numerical and –Speed ability domain did not correspond to the already established faceted model of social intelligence. Therefore, the model of reference only relied on the BIS-Reasoning and BIS-Memory cells, cross-classified with Verbal and Figural-Spatial abilities. The model of reference for the social intelligence domain relied on four combined indicators for the cross-classification of social understanding and memory as operative factors and language-based and language-free contents (i.e., SUl, SUlf, SMl, SMlf; with l = language-based and lf = language-free contents). Thus, the operative and the content structure corresponded with the respective structure of the BIS.

The first model to investigate a combined faceted structure integrated social intelligence completely into the BIS structure (Model P). For example, the language-based social understanding cell (SUl) loaded on the BIS-Reasoning and the BIS-Verbal factor, and the language-free social memory cell (SMlf) loaded on the BIS-Memory and the BIS-Figural factor. Model P showed good data fit. The factor loadings on the content factors, however, varied substantially between the social and academic intelligence cells with mostly larger loadings of the BIS-cells. Again, two equality constraints had to be introduced because error terms appeared at the lower bound. The next model (Model Q) established a separate "social content" factor with loadings of all the social intelligence cells which no longer were allowed to load on the BIS content factors. It was still necessary to introduce two equality constraints on error terms. The model showed very good fit statistics (CFI = .996; χ^2 = 11.452, p=.323). Models P and Q were nested. The χ^2-differences test showed a significantly better data fit for Model Q with the additional social content factor (χ^2-difference = 8.685, df = 2, p < .01). Figure 8.13 presents the standardized solution of Model Q. The loading pattern on the content-factors was still heterogeneous, particularly on the social content factor. Moreover, the social and academic intelligence cells loaded differentially on the reasoning factor. The operative factors were not meaningfully intercorrelated (*r* = .08).

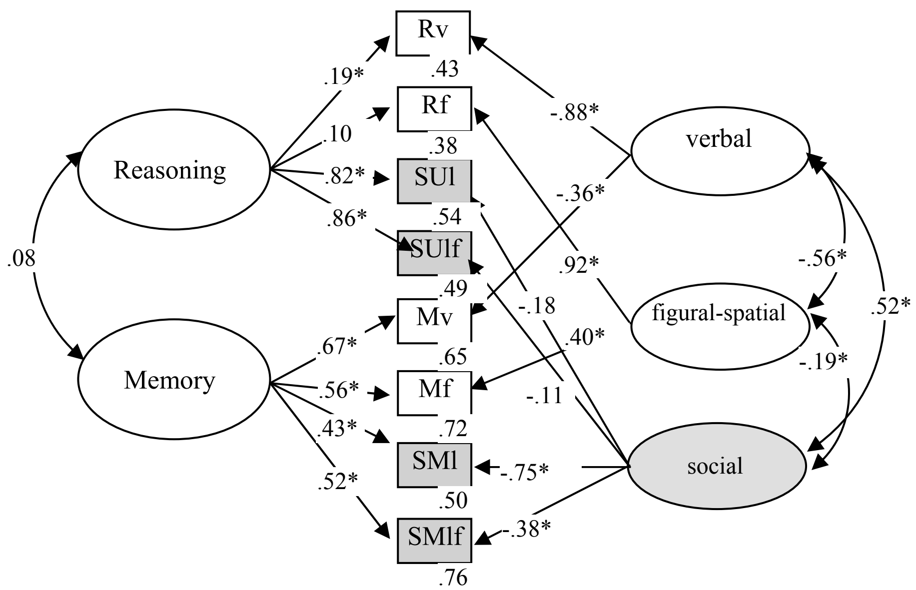


Figure 8.13

Standardized Solution of Combined Faceted Model of Social and Academic Intelligence (Model Q) (Study 2)

Note. CFI = .996; χ^2 = 11.452, p=.323; * p < .05; error terms are enclosed with the manifest variables
R = BIS-Reasoning, M = BIS-Memory, v = BIS-Verbal contents, f = BIS-Figural-spatial contents, SU = social understanding, SM = social memory, l = language-based, lf = language-free contents

A third model postulated an additional "social operative" factor with loadings of all social intelligence cells (Model R), and omitted the separate "social content" factor. These cells loaded again on the content factors of the BIS-Model as postulated in Model P. The model fit, however, was worse than for Model Q (CFI = .966; χ^2 = 23.253, p < .01) and the model also needed equality constraints.

The results from this analysis pointed towards a possible classification of the social intelligence operative structure into the BIS operations whereas the social intelligence contents could not be subsumed under the BIS-content structure. However, the loading pattern was rather heterogeneous and some psychometric problems occurred so that equality constraints needed to be introduced.

Relationship to Academic Intelligence (Structural Independency) – Research Question 2C2

As a last step, the structural independency of social from academic intelligence should be investigated. Therefore, the BIS variance was partialled out of the social intelligence variables by entering the BIS scales (i.e., the broad operative and content ability factors) into a regression on every single social intelligence variable and saving the standardized residuals. Then, two further models were established that had shown good data fit in the analysis of the internal structure of social intelligence (Chapter 8.4.3.1). First, the structural model was replicated with a social understanding and a social memory factor (Model S equivalent to Model E). The model showed very good data fit (CFI = 1.000; χ^2 = 32.241; p=.554). No problems were encountered during optimization. The loading pattern was rather homogeneous and the two ability factors were small but meaningfully intercorrelated. Moreover, the correlation between the ability factors in Model S (r = .19) was not substantially smaller than that in the original Model E (r = .20). Figure 8.14 presents the standardized solution of Model S. Furthermore, the faceted model of social intelligence with two correlated operative and two correlated content factors should be replicated by examining the residuals (Model T with an equivalent Model I). This model also showed very good fit statistics although again, two equality constraints needed to be introduced because of error terms at the lower bound. A comparison of the nested Models S and T showed no significant difference and the more parsimonious Model S was accepted (χ^2-difference = 10.205, df = 9, n.s.). In summary, this analysis showed that the structure of social intelligence was independent from the common BIS variance.

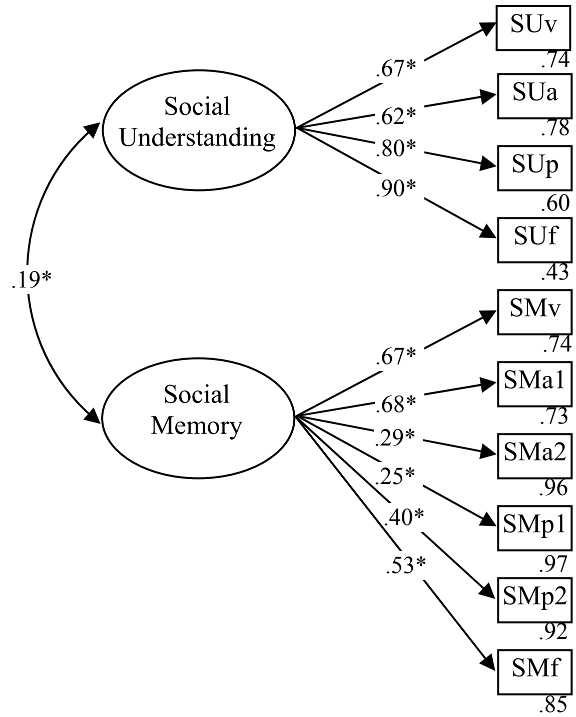


Figure 8.14

Standardized Solution of the Structural Model of Social Intelligence (Model S) when Variance of the BIS was Controlled for (Study 2)

Note. CFI = 1.000; χ^2 = 32.241 (p=.554); * p < .05; error terms are enclosed with the manifest variables
SU = social understanding, SM = social memory, v = written language, a = spoken language, p = pictures, f = videos

Relationship to Personality Traits – Research Question 2C1

To prove the divergent construct validity, social intelligence should also be discriminable from personality traits. Several trait inventories were applied in the present study (i.e., the Big Five assessed by the NEO-FFI, two subscales of Empathy: empathic compassion and cognitive perspective taking, Altruism, and Depression). Table 8.20 presents the correlations between the trait variables and the social intelligence tasks. Only the social understanding tasks showed unsystematic correlations with some personality traits. The scales SUv and SUa were negatively intercorrelated with Agreeableness and Conscientiousness, SUv negatively with empathic compassion, SUa negatively with Extraversion and positively with Depression. The scale SUp correlated slightly positively with Neuroticism and negatively with Agreeableness. However, the correlative pattern showed that the overlap between the social intelligence tasks and personality traits was generally small and above all smaller than the within-domain correlations. It was not clear, however, why the scale intercorrelations were partly substantially negative.

Table 8.20

Correlations of Social Intelligence Tasks with Personality Traits (Study 2)

	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C	EC	PT	Altru -ism	Depression
SUv (182)	.055	-.130	-.005	-.284**	-.165*	-.203**	-.093	.071	.064
SUa (182)	.131	-.204**	-.042	-.243**	-.212**	-.063	.013	-.070	.238**
SUp (182)	.149*	-.087	-.011	-.149*	-.140	-.024	.056	.083	.144
SUf (182)	-.003	-.014	-.065	-.151*	-.107	-.071	.007	.142	.051
SMv (182)	.089	-.036	-.033	.033	-.107	.019	-.062	.008	.056
SMa1 (182)	.077	-.108	.010	.062	-.110	-.034	-.003	.005	.086
SMa2 (182)	-.042	.011	.037	.074	.056	.038	.084	.005	-.088
SMp1 (177)	-.119	.051	.090	.091	-.002	-.014	.069	.000	-.075
SMp2 (177)	.040	-.060	-.038	.041	.047	-.009	-.053	.113	.008
SMf (179)	.145	-.063	-.080	.126	-.086	-.087	.019	.017	.011
SPv1 (182)	-.043	-.127	.020	-.061	.136	-.044	.069	-.106	.012
SPv2 (182)	-.064	-.024	-.094	-.022	.101	-.077	.123	-.106	-.080
SPa1 (182)	.074	.010	-.169*	.070	-.043	-.049	-.061	.011	-.029
SPa2 (181)	.038	.044	-.065	-.071	.071	.065	.048	.060	-.012
SPp1 (182)	.028	-.058	.013	-.111	.029	.082	.019	.000	.039
SPp2 (182)	.065	-.132	.061	.049	.073	.066	.110	.019	.109
SPf1 (177)	.073	.001	-.004	-.059	.021	-.065	.102	.025	.096
SPf2 (182)	.142	-.057	-.062	.035	-.065	-.101	-.020	-.025	.153*

Note. pairwise N in parentheses, * p < .05, ** p < .01
SU = social understanding, SM = social memory, SP = social perception, v = written language, a = spoken language, p = pictures, f = videos, NEO-N = Neuroticism, NEO-E = Extraversion, NEO-O = Openness, NEO-A = Agreeableness, NEO-C = Conscientiousness, EC = empathic compassion, PT = perspective taking

Appendix H presents the intercorrelations within the personality traits that were generally high and all in the expected direction so that no syntax mistake could be responsible for the negative scale intercorrelations with social understanding.

8.4.4 Further Exploratory Analyses

8.4.4.1 Relationship to Self-Report Data – Research Question 3A

Two self-report questionnaires of social and emotional skills were assessed in the present study, the Social Behavior Questionnaire of Amelang et al. (1989) and the Schutte Emotional Intelligence Scale (Schutte et al., 1998). The two measures were substantially intercorrelated ($r = .509$; N = 182; $p < .001$). The social intelligence tasks were generally uncorrelated with the self-report inventories. Correlations ranged between $r = -.144$ and $r = .108$ with the Social Behavior Questionnaire and between $r = -.182$ and $r = .118$ with the Schutte Emotional Intelligence Scale. These correlations were reduced to zero when variance of the personality traits was controlled for.

This latter finding was due to a large overlap between the self-report inventories and the personality traits presented in Table 8.21. Entering all personality traits as predictors into a regression analysis with each self-report inventory as criterion measures, the amount of explained variance was rather large. Self-reported social behavior could be explained with $R^2 = .282$ ($F_{9/172} = 7.517$; $p < .001$), and self-reported emotional intelligence with $R^2 = .443$ ($F_{9/172} = 15.184$; $p < .001$). The scale intercorrelation based on the standardized residuals was reduced to a meaningful extent compared to that of the original scales. However, they were still substantially correlated ($r = .294$; N = 182; $p < .001$) which pointed to some meaningful construct overlap of self-reported social and emotional skills.

Table 8.21

Correlations of Self-Report Inventories with Personality Traits (Study 2)

	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C	Altru-ism	EC	PT	De-pression
SB Questionnaire	-.265**	.344**	.251**	.097	.255**	.309**	.153*	.248**	-.291**
Schutte EI Scale	-.459**	.456**	.164*	.194**	.440**	.217**	.175*	.330**	-.500**

Note. N = 182; * p < .05; ** p < .01
SB = Social Behavior; EI = Emotional Intelligence, NEO-N = Neuroticism, NEO-E = Extraversion, NEO-O = Openness, NEO-A = Agreeableness,
NEO-C = Conscientiousness, EC = empathic compassion, PT = perspective taking

8.4.4.2 Gender Differences

Overall Gender Differences of Social Intelligence Tasks – Research Question 3B

The question of gender differences in the social intelligence tasks was only exploratory. As was already mentioned in Study 1, some authors interpreted gender differences in favor of women as a positive validity result for measures of social and emotional abilities (Hall, 2001; Schutte et al., 1998). It was not self-evident to formulate such an assumption based on the present conceptualization of social intelligence. Therefore, the alpha-level was adjusted in order to account for the accumulation of error (new alpha-level = .003). Applying this alpha-level, three social memory tasks and one social perception task showed significant gender differences in favor of women (SMv: t =-3.979, df = 180, p < .001, Cohen's d = .573; SMa1: t = -4.031, df = 180, p < .001, Cohen's d = .585; SMf: t = -3.233, df = 177, p=.001, Cohen's d = .478; SPv2: t = 3.395, df = 180, p < .001, Cohen's d = .497). Nearly all remaining tasks showed gender differences in the same direction, but did not reach significance. Interestingly, three social understanding tasks (SUv, SUa, and the personality ratings) showed a nonsignificant tendency of males to perform better.

Gender Differences in Targets, Subjects, and Their Interaction – Research Question 3B1

The more interesting question referred to the effect of the target gender on the performance of male and female subjects. Therefore, a general scale of every target was composed by aggregating all of the items from the final social understanding scales across the four content scales and the personality rating scale. All general scores of male and female targets were combined to one score for each target gender. These scores were highly correlated (r = .826, p < .001, N=182). In Study 1, male performances in judging male and female targets were positively intercorrelated as were female performances in judging male and female targets. This result was replicated in the present study which included two additional male and female targets. Results showed substantially positive correlations for both gender groups with r = .849 (p < .001; N = 75) of males judging males, with males judging females and r = .810 (p < .001; N = 107) of females judging males with females judging females. No interaction effect occurred that could have been attributed to an influence of the similarity of judge and target in terms of gender.

To investigate the effect of the target gender interaction with the subjects' gender more closely, a two-factor analysis of variance (repeated measurements) was conducted postulating one between-subjects factor (i.e., the subject gender) and one repeated measurement factor (target gender). Results are presented in Table 8.22 and contradict the results from Study 1. Only the main effect of the repeated measurement reached significance showing a lower item difficulty for male targets than for female targets. No significant interaction effect was discovered. Figure 8.15 illustrates the results by plotting the group means.

Table 8.22

Results of Two-Factor Analysis of Variance of the Effect of Target and Subject Gender on Performance in Social Understanding Scales (Study 2)

Source of variance	Sums of square	DF	Mean squares	F	Probability	Eta²
Repeated measures (target gender)	4.489	1	4.489	210.599**	<.001	.093
Between-subjects (subject gender)	.026	1	.026	.115	.735	
Interaction effect	.000	1	.000	.006	.940	

Note. N = 182

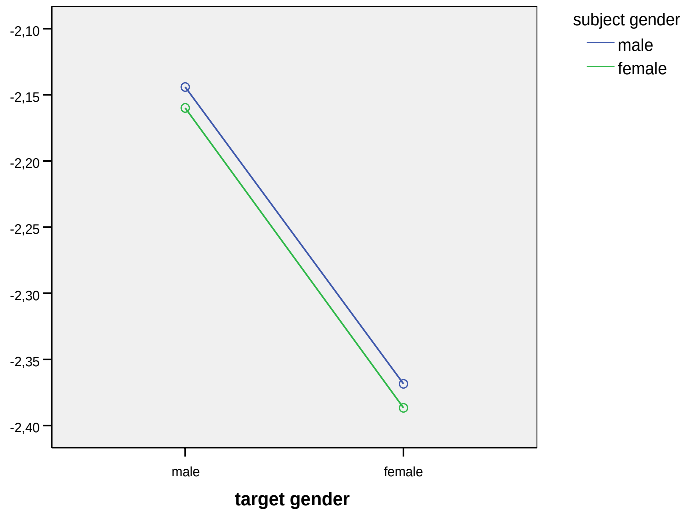


Figure 8.15

Results of the ANOVA Showing Gender Effects on Performance in Social Understanding Scales Based on all Targets (Study 2)

In Study 1, female targets had been easier to judge. The present finding was all the more surprising because four of the targets were the same as in Study 1. In order to look at this more closely, Table 8.23 presents the means and standard deviations of the single targets comparing Study 1 and 2. It must be noted that the scale range was changed from Study 1 to Study 2 which is indicated in the right column of the table. The maximum possible deviation is “-5” in Study 1 and “-6” in Study 2. It turned out that the two female targets added in Study 2 (Conny (CK), and Hannah (HR)) were the scenarios with the highest difficulty. Moreover, the scenario with target Katharina (KL), which was among the easiest in Study 1, turned out to take the third rank in terms of difficulty after CK and HR.

Table 8.23

Means and Standard Deviations of Social Understanding Performance for the Single Targets (Study 2)

Target	RF	BS	CK	CP	KL	FB	HR	MM	scale
Gender	female	male	female	male	female	male	female	male	
Study 1	-1.269 (.488)	--	--	-1.444 (.433)	-1.340 (.341)	--	--	-1.555 (.373)	-5 – 0
Study 2	-2.038 (.368)	-2.142 (.398)	-2.425 (.448)	-2.073 (.396)	-2.362 (.406)	-2.143 (.475)	-2.690 (.558)	-2.256 (.435)	-6 – 0

Note. *sd* in parentheses

Thus, the contradicting results did not only seem to be due to the adding of four different and more difficult scenarios, but also to a change in difficulty of the scenario Katharina (KL).

To confirm this finding, the two-factor analysis of variance was conducted again utilizing only the four targets that were applied in Study 1 to the present sample. Results showed a significant effect of the repeated measures related to the target gender ($F_{1/180} = 4.675$; $p < .05$) and a significant interaction effect of target and subject gender ($F_{1/180} = 8.256$; $p < .01$). Figure 8.16 plots the group means and demonstrates the diverse effect compared to Study 1 (see the small graph on the right side showing Figure 7.12 from Study 1). The performance in scenarios relying on female targets was equal to that of Study 1 with, however, only a small tendency for females to perform better. A diverse effect occurred for male targets with a substantially better performance by male subjects. It could not be determined at this point whether the task modifications contributed to this change or whether the sample in the present study found it more difficult to judge, for example, Katharina. A post-hoc explanation could account for the university background of the scenario Katharina, so that the student subjects in Study 1 could judge Katharina's mental states better than the heterogeneous sample in Study 2. However, the effect was obscured by the use of different items in both samples and across the different scenarios.

In any case, these results still needed replication, possibly utilizing even more targets and a balanced ad heterogeneous sample in terms of similar educational and occupational backgrounds.

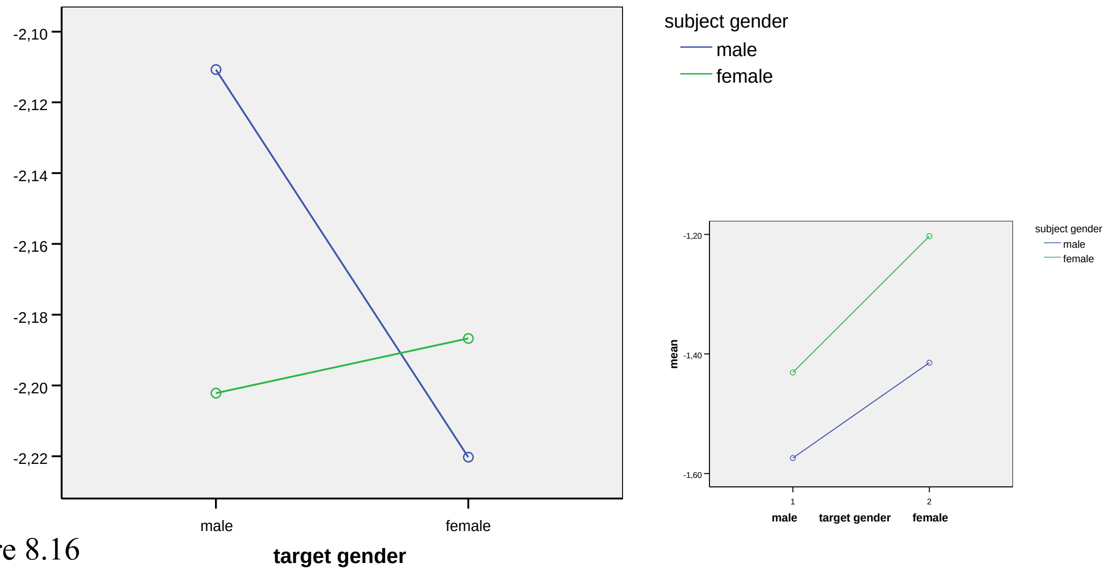


Figure 8.16

Results of the ANOVA Showing Gender Effects on Performance in Social Understanding Scales Based on Four Targets Applied in Study 1 (Study 2)

Note. Results from Study 1 on the right side

8.4.4.3 Scoring Alternatives – Research Question 3C

Group Consensus Scoring – Research Question 3C1

Group consensus scoring (i.e., proportion scoring) assigns the relative frequency of a rating category in the sample to a subject's answer matching this rating category. This method is frequently applied as an alternative scoring procedure when no objectively correct answer can be identified. In some approaches, group consensus scoring seems to be selected because it represents a convenient procedure requiring no additional effort towards scale construction (e.g., the Faces Test of the MSCEIT). However, several researchers have questioned the adequacy of this scoring method (Matthews et al., 2005; Schulze et al., 2007; Tagiuri, 1969). The most frequent critique addresses the problem of item difficulty. One criticism points out that single items cannot identify highly able subjects (whose answers are rare compared to the standard answer of the sample average) because consensus scoring favors agreement with the sample average (see Chapter 5.2.4.2 for a more detailed description of the entire procedure and the associated discussion). Thus, the concept of item difficulty cannot be applied to items scored by group consensus. It could be argued, however, that this ceases to be problematic when items are aggregated to a scale. But empirical studies have not addressed this problem so far. In this respect, there are other approaches that investigated the relationship of group to expert consensus scoring as one possible external standard to evaluate the usefulness of this scoring procedure. The scale intercorrelations were rather low in the scales of the predecessor of the MSCEIT (i.e., the MEIS; r = .26) and rather high for the MSCEIT scales (r = .96). Both correlations were critized for either being too low or too high (Roberts, et al., 2001). Consequently, it had to be questioned what the “adequate” correlation size would to be expected when group consensus scoring is correlated with other scoring procedures.

The present study, however, applied target scoring. The question addressed in the present analysis referred to whether correlations between group consensus scores and target scores of the same scale are generally capable of providing useful information about either scoring method. In this respect, target scoring could as well represent any other scoring method applying comparable scoring algorithms (e.g., specific expert scoring procedures). Thereby, a more thorough look could be given to the effect of item difficulty on the intercorrelations between different scoring methods.

In Chapter 5.2.4, an artificial simulation was undertaken by varying the difficulty of items scored by target scoring and assessing the effect on the bivariate distributions between the two scoring methods. In the present analysis, the same question was addressed relying on the data of the social understanding tasks in the present study. The two scoring methods were shown to be highly correlated on the general level of the content-related scales (r = .687 - .783 for the spoken language, pictorial, and video-based scales; see Table 8.12 for the entire correlation matrix) with smaller correlations for the written language scales and the personality rating scales (*r* = .314 / .466 for SUv and SUps respectively). The first three data columns in Table 8.24 summarize these parameters showing the correlations between the scoring methods of each content-related scale, the respective reliability coefficients, means, and standard deviations. It could be speculate that the lower correlation between the written language scales were attributable to the enhanced mean difficulty of the target scoring of about one standard deviation (*m* = -2.628 compared to a mean difficulty of -2.213 for the remaining scales). It was worth noting, however, that the mean target score of the personality rating scale was at about the same level as those scales (SUa, SUp, and SUf) that showed the highest correlations between the scoring methods.

Table 8.24

Scale and Item Properties of Social Understanding Scales Separately for Different Scoring Procedures and Item Difficulties

SU scale	$r_{(TS, GCS)}$	α TS	α GCS	Mean (TS)*	Mean (GCS)*	Deviation Target Answer – Sample Mode**			
						0	-1	-2	(-3) – (-6)
SUv	.314	.752	.765	-2.628 (.397)	.231 (.023)	6	7	7	16
SUa	.743	.791	.839	-2.269 (.423)	.244 (.024)	18	13	11	3
SUp	.687	.754	.792	-2.252 (.422)	.232 (.023)	17	16	6	10
SUf	.783	.841	.834	-2.122 (.389)	.238 (.024	25	26	9	9
SUps	.466	.851	.831	-2.208 (.472)	.263 (.033)	13	27	18	2

Note. * *sd* in parentheses; **number of items
TS = target scoring, GCS = group consensus scoring, SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings

This contradictory finding was further examined. The focus therefore switched from the scale to the item level. Items were classified according to the deviation between the target answer of an item and the mode of the sample's raw answers. The target answer represented the rating category that provided the highest (best) score in target scoring when selected by a subject. At the same time, the mode of the sample's raw answers represented the rating category providing the highest (best) score in group consensus scoring. Thus, a "0" indicated zero deviation between the target answer and the mode, "-5" a deviation of five points between the target answer and the sample mode. The right columns of Table 8.24 present the number of items within each scale, classified into different deviation categories between the target answers and the mode value of the sample's raw answers. The full range of possible deviations is not displayed because of very few answers in the categories with the largest deviations. The categories "-3" to "-5" were combined. A deviation of "-6" did not occur. Obviously, the distribution of different deviation categories differed between the content-related scales with a strikingly different distribution for the written language scale. Additionally, the personality rating scale stood out because, as opposed to the remaining scales, a zero deviation occurred substantially less than a deviation of "-1".

In order to analyze the effect of item difficulty on the inter-scoring correlations more closely based on statistical methods, the two extreme deviation classes were first inspected.

Therefore, for the two example scales SUf and SUps, items were aggregated to new compound scores. Thus, the "low difficulty scale" only contained items where the sample's mode answer was identical with the target answer (i.e., a deviation of "0"). The "high difficulty scale" only contained items with a large deviation ("-3" to "-5"). The same items scored by group consensus scoring were aggregated to the respective scales. Prior to any further analysis, the reliabilities of the resulting four scales per content domain were calculated. These were rather low due to a reduction in the number of items. The reliabilities were estimated for a test length of 60 items by the Spearman-Brown-Formula and turned out as sufficient (i.e., the number corresponded to the mean number of items in the original content-related scales).

Afterwards, the correlations between the scoring methods were investigated separately for the two content scales and the item classes related to the deviation categories (Table 8.25). The correlations between the scoring methods were consistent for each of the content domains. The inter-scoring correlations for the "low difficulty scales" were *r* = .766 / .787, respectively with SUf and SUps. The inter-scoring correlations for the "high difficulty scales" were *r* = -.186 / -.226, respectively.

Table 8.25

Example Scales: Effect of Item Difficulty on the Correlations Between the Scoring Methods

	Low difficulty: Zero deviation					**High difficulty: Deviation of "-3" to "-5"**				
SU scale	$r_{(TS, GCS)}$	**α TS****	**α GCS****	**Mean (TS)***	**Mean (GCS)***	$r_{(TS, GCS)}$	**α TS**	**α GCS**	**Mean (TS)***	**Mean (GCS)***
SUf	.905	.766 .577 (25)	.731 .531 (25)	-1.374 (.385)	.256 (.033)	-.186	.755 .316 (9)	.548 .154 (9)	-3.287 (.574)	.221 (.033)
SUps	.850	.787 .444 (13)	.830 .514 (13)	-1.377 (.381)	.272 (.046)	-.226	.669 .063 (2)	.779 .105 (2)	-2.993 (1.052)	.264 (.093)

Note. N = 182; * *sd* in parentheses; ** number of items in parentheses
upper line of reliability parameters indicates reliability estimation based on Spearman-Brown-Formula (60 items)
TS = target scoring, GCS = group consensus scoring, SU = social understanding, f = videos, ps = personality ratings

It could be reasoned that the two scoring methods for the "high difficulty scales" were not intercorrelated because these items measured not the same measurement construct. However, the intra-scoring correlations suggested that the same underlying ability was measured by both, the "high" and the "low difficulty scales". The empirically determined correlations were corrected for attenuation because of the low reliabilities of the original scales. The intra-scoring correlations of the SUf scales between the scales based on different item difficulty were estimated at r_{corr} = .794 (r_{emp} = .338) for target scoring and for r_{corr} = .980 (r_{emp} = .279) for group consensus scoring. The correlations between the SUps scales of item groups were estimated at r_{corr} = .120 (r_{emp} = .020) for target scoring and at at r_{corr} = .892 (r_{emp} = .207) for group consensus scoring. The low correlation between the “high difficulty” personality rating scales for target scoring was based only on two items which restricted the validity of the scale. But in general, the results supported the original assumption that the inter-scoring correlations were largely influenced be item difficulty, at the same time measuring the same construct.

Table 8.25 also shows that the expected large difference in mean difficulty of the target scoring scales was not reflected in the scales based on group consensus scoring. Interestingly, the scales based on the two difficulty classes in the target scoring scales showed a standardized mean difference of Cohen’s d = 4.913 / 3.424 (respective for SUf and SUps). The scales scored by group consensus showed only a standardized mean difference of Cohen’s d = 1.417 / .242 (respective for SUf and SUps). This result showed that the means in group consensus scoring scales did not seem to be affected by any external standard such as a target answer.

So far, two extremes of item difficulties were regarded showing two extremes in correlation size between the scoring methods. As a last step in this analysis, the video-based scale was picked exemplarily and items with a deviation of “-1” between the target answer and the sample’s mode answer (i.e., “medium difficulty scale”) were selected. This represented the category most frequently occurring within the present data. The overall correlation between the scoring methods was r = .772 (Cronbach’s alpha =.717 / .605 respectively for target and group consensus scoring, based on 26 items). The simulation in Chapter 5.2.4 had shown that a curvilinear relationship appeared under the condition of at least one point deviation between the target score and the sample’s mode answer and the same target raw answer. Conventionally, within one aggregated scale, different item difficulties and items based on different target answers are desired and combined to a composite score. The curvilinear relationship, however, can also be shown in the present data when single items are considered.

Only those items were selected out of the video-based scale, all showing a deviation of “-1” between the target answer and the sample’s mode answer and all relying on a target answer of “6”. The inter-scoring correlation was *r* = .446. Cronbach’s alpha for this scale was .457 for target scoring and .410 for group consensus scoring. Figure 8.17 presents the bivariate distributions between the target and the group consensus scores for all of the aforementioned video-based scales showing *(a)* the “low difficulty scale” (*r* = .905), *(b)* the “high difficulty scale” (*r* = -.186), (c) the “medium difficulty scale” (i.e., only items based on a deviation of -1; *r* = .772), and (d) a selection of items from the “medium difficulty scale” based only on target answers of “6” (*r* = .446).

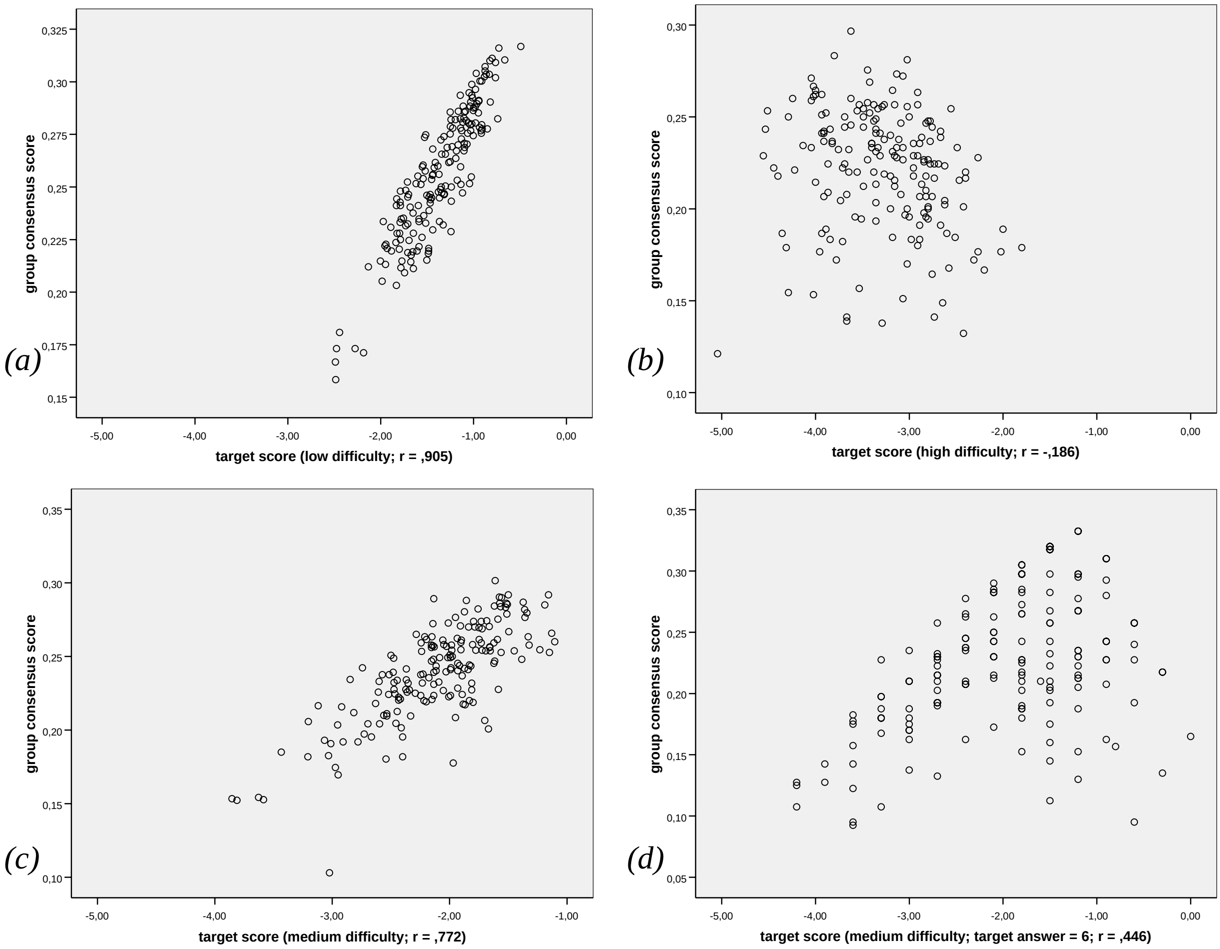


Figure 8.17

Bivariate Distributions Between Target and Group Consensus Scores Based on Different Item Difficulties

The Figure clearly shows that for single classes of items, the shape of the distribution depended on the item difficulty *and* that a curvilinear relationship emerged when focusing on items that relied on the same target answer. The shape of distributions were similar when items were regarded which, for example, relied on a target answer of “5” and showed a deviation of “-2”.

Conclusively, it appeared problematic to interpret the correlations of group consensus scoring scales with other scoring methods providing a criterion such as target scoring procedures or types of expert scoring procedures. Obviously, the correlations are determined by several different classes of item difficulties so that the aggregation of items of different difficulties obscures the true relationships between the scoring methods.

Correlations-Based Scoring – Research Question 3C2

The correlations-based scoring method is conventionally applied in interaction research and is a subtype of target scoring (Bronfenbrenner et al., 1958; Funder, 2001; Ickes et al., 1990; Kenny & Winquist, 2001; Snodgrass, 2001). Correlations-based scoring assigns each subject (i.e., the judge) a score based on the correlation between the subject's and the target's answers on a set of items. Therewith, the information from the single items is lost. Interaction paradigms typically involve the same person as both, the target and the judge, being put into an encounter. The scoring method is sought to account for an interaction effect of the accurate sending and perceiving of cues (Ickes et al., 1990; Snodgrass, 2001). Snodgrass (2001) provided a detailed description of this scoring method and stressed both, the limitations and advantages. She acknowledged that other scoring methods were much more fine-grained and would show report higher reliabilities. By correlating a set of items in order to build the score, a subsequent reliability analysis could only rely on the set information and was only possible if several targets were applied. Snodgrass, however, did not report reliability coefficients. According to Snodgrass, correlations-based scoring was influenced by stereotypes which she equaled with "good guesses" or the use of heuristics in social judgments. She further claimed that correlations-based scoring compensated for different rating tendencies of the judge and the target by relying on z-standardized scores. For example, if a target person tended towards extreme answers and a judge towards the middle of the scale, this would be equaled with inaccuracy in the target scoring procedure applied in the present study. The overall target answers were only inspected for tendencies and did not show specific features. There could be, however, rating tendencies of single subjects or for subsets of items so that it seemed interesting to look at this type of scoring procedure.

In the present analysis, correlations-based scoring was applied to the content-related scales of the social understanding tasks. Therefore, each subject's answers were correlated with the target answers for the item sets of one content domain. The correlations were transformed to Fisher's Z to allow a comparison between them. Every subject received eight scores for each content-related scale representing the correlations with the answers of the eight targets. The correlations relied on a number of items between nine and 23 (Snodgrass reported a basis of 13 items for calculating the correlations). Thus, it was possible to conduct further analysis separately for the content domains which would not have been possible when only the correlation across all items had been calculated.

A reliability analysis of the content-related scales (each consisting of eight scores associated with the eight targets) showed generally low reliability coefficients. The written language scale even showed a negative value due to a negative covariance matrix between the items. Cronbach's alpha for the content scales and the personality rating scale was -.009 / .295 / .058 / .294 / .215, respectively for SUv / SUa / SUp / SUf / SUps. However, since Snodgrass did not report reliabilities, no comparison was possible. It could be speculated, however, that this type of scoring produced homogenous variance *within* one scenario instead of within the content-related scales *across* the different scenarios (i.e., attached to one target person). Therefore, the reliability of the target-related general scales across the contents was also analyzed. The reliabilities were low on average with only one exception of the scenario Renate (RF) with $\alpha = .423$. The remaining coefficients were $\alpha = .027$ / .104 / .002 / .257 / .111 / .128 / -.018 for the other seven scenarios (BS, CK, CP, KL, FB, HR, and MM respectively). Thus, the content-related scales were applied hereafter, acknowledging that it was problematic to aggregate the scales due to the low reliabilities.

The descriptives of the scales and the correlations within the social understanding tasks are presented in Table 8.26. All scales were normally distributed, the distribution parameters did not indicate substantial deviations from the normal distributions (i.e., skewness between -.241 and .113; kurtosis between -.299 and .734. No outlying values were discovered. The range of scores was rather large. The smallest range was discovered for the written language scale (scores between .11 and .54), the largest range for the personality rating scale was (scores between -.14 and .73). According to Kenny and Winquist (2001), this large variance was a precondition for detecting covariances with other variables and it showed that the answers across subjects did not converge too much. The correlations within the social understanding scales, however, did not support a coherent domain of social understanding and thus no convergent validity. Only three correlations reached significance (i.e., between SUv and SUp, and SUf with SUa and SUp). The correlations with the scales based on target scoring were inconsistent with the largest correlations between the scales of the same content domain ($r = .271 - .430$).

Table 8.26

Descriptives of Social Understanding Scales (Correlations-Based Scoring), Intra- and Interscoring Correlations

			CBS				TS				
CBS	**M**	**SD**	**SUv**	**SUa**	**SUp**	**SUf**	**SUv**	**SUa**	**SUp**	**SUf**	**SUps**
SUv	.340	.089					.404**	-.003	.227**	.182*	.162*
SUa	.435	.098	.053				.066	.430**	.201**	.154*	.014
SUp	.301	.104	.250**	.112			-.050	-.130	.279**	.039	-.162*
SUf	.370	.135	.131	.218**	.261**		.001	.055	.243**	.382**	-.125
SUps	.317	.140	.144	.069	.001	-.061	.106	-.049	.035	-.028	.271**

Note. N = 182, * p < .05, ** p < .01
SU = social understanding, v = written language, a = spoken language, p = pictures, f = videos, ps = personality ratings, CBS = correlations-based scoring, TS = target scoring

As a next step, the correlations of the correlations-based scales with the remaining social intelligence tasks and with measures of academic intelligence and personality were explored. The correlations with the social memory tasks still showed rather low but positive correlations. The scale SUa, however, correlated more highly with the social memory tasks; *r* = .227 / .270 / .338, for SMf, SMv and SMa1 respectively (see Table 8.12 for the original correlation matrix). The correlations with the social perception tasks as well as with the BIS cells did not change meaningfully. In contrast, the correlations with personality traits were substantially different (see Table 8.27).

Table 8.27

Correlations of Social Understanding Tasks (Correlations-Based Scoring) with Personality Traits

	NEO-N	**NEO-E**	**NEO-O**	**NEO-A**	**NEO-C**	**EC**	**PT**	**Altruism**	**Depression**
SUv	-.001	-.025	.036	-.073	.002	-.050	-.030	.061	-.053
SUa	.039	-.082	-.018	.022	-.017	.038	.109	-.031	.035
SUp	-.061	.151*	-.080	.172*	.149*	.059	.124	.095	-.125
SUf	-.148*	.181*	.058	.162*	.201**	.162*	.111	.244**	-.162*
SUps	-.029	.030	.085	.008	-.118	-.007	.024	-.086	.039

Note. N = 182; * p < .05; ** p < .01
SU = social understanding, , v = verbal, a = auditory, p = pictorial, f = video-based, ps = personality ratings; NEO-N = Neuroticism, NEO-E = Extraversion, NEO-O = Openness, NEO-A = Agreeableness, NEO-C = Conscientiousness, EC = empathic compassion, PT = perspective taking

The target scoring scales had shown partly negative correlations with, for example, Extraversion, Agreeableness or Altruism (see Table 8.20 for the original correlations). The newly scored scales correlated to a generally small extent with the personality traits which still supported the divergent construct validity. The sign, however, changed for most of the correlations, particularly for the video-based scale. Although no real expectations were formulated for the relationship of the social intelligence tasks to personality traits, the positive correlations conformed with what is reported in the academic intelligence literature (Ackerman & Heggestad, 1997) and with what could be expected for social or emotional abilities within a social competence framework (see Chapter 3 and Figure 3.1). The correlations with self-reported social and emotional skills were still around zero with one significantly positive correlation between the video-based scale and the Schutte Emotional Intelligence Scale of $r = .208$ (N = 182; $p < .01$). The correlations with the PONS did not change meaningfully.

As a last step, gender differences were explored. In contrast to the target scoring scales, all variables showed female subjects perform better with two significant mean differences of the pictorial and video-based scales ($t = 3.637$ / 2.697; $df = 180$; $p < .01$; $d = .529$ / .400; respective for SUp and SUf).

In summary, the social understanding scales scored by correlations-based scoring showed meaningful divergence to the target scoring scales. The problems, however, concerned the low reliabilities and the loss of information of single items. Moreover, the correlations between the social understanding scales did no longer support a coherent ability domain. These problems notwithstanding, the lack of negative correlations with personality traits such as Agreeableness and the gender differences in favor of women conformed with the literature. At present, it could not be clarified whether the change in these results was only due to the purported compensation of different rating tendencies or to the different scoring procedure in general. In general, there appeared no reason for substituting the target scoring procedure by the correlations-based scores.

8.4.4.4 Exploring the Dimensionality of the Target Score: Effect of Assumed and Real Similarity – Research Question 3D

Effect of Similarity on Social Understanding Scales Scored by Traditional Target Scoring

The idea underlying the analysis described below resulted from Cronbach's critique of the traditional target scoring procedure (Cronbach, 1955). He claimed that the target score consists of various components representing different types of accuracy (e.g., stereotype or differential accuracy; see Chapter 5.2.4.1; for a detailed description see also Kenny & Winquist, 2001). These accuracy components are sought to be influenced by the assumed and the real similarity between judge and target. Cronbach established a complicated scoring system that purportedly allowed the identification of the separate variance components. This scoring system, however, was criticized for being too analytical (Funder, 2001). Moreover, it only works when all targets are judged on the same items and so, it could not be applied to the social understanding scales in the present study.

Therefore, the current section attempted to examine the effect of the operationalized assumed and real similarity between judge and target on the accuracy of the social understanding tasks. It had to be accounted for that the original idea of Cronbach only referred to the similarity in the use of rating formats. However, as several researchers acknowledged; the judge-target relationship, in terms of familiarity or similarity, effects the accuracy of judgments (Bronfenbrenner et al., 1958). Thus, it appeared reasonable to suggest that the interaction of assumed and real similarity in terms of the biographical data and personality as coded here, could determine the judgmental accuracy to a certain extent. Since the similarity concept was related to each target person separately, the analysis relied on the general performance in one scenario across the content domains. Scores were built including the items from the content-related scales. The reliability coefficients were found to be reasonable. Cronbach's alpha for the general scores were .674 / .582 / .711 / .628 / .632 / .774 / .845 / .615 for the scenarios in the order of testing (i.e., RF, BS, CK, CP, KL, FB, HR, and MM). The scenarios were highly intercorrelated with *r* between .416 and .646.

The indicator for the assumed similarity was assessed at the end of each scenario (i.e., "How similar do you think you are to the target person?", 7-point rating scales from "*not at all similar*" to "*extremely similar*"). Additionally, several indicators of the real similarity between the subjects and the targets were built by comparing different biographical data (i.e., gender, age, parenthood, and education) and by correlating the NEO-FFI profiles. The indicators for the biographical data were dichotomously scored with "0" (dissimilar) and "1" (similar). For example, if subject and target were similar in parenthood (i.e., both having children), the subject received a score of "1". For assessing the similarity in age, different age groups were composed around the target's age that received different degrees of similarity (i.e., within +/- 2 years meant a score of "4", within +/- 3 through 6 years meant a score of "3", within +/- 7 through 10 years meant a score of "2", and beyond +/- 11 meant a score of "1"). Thus, both dichotomous (i.e., similarity in gender, parenthood, and education) and continuous variables (i.e., assumed similarity, similarity in age, the correlation with the personality profile of each target) were applied in the following analysis.

Prior to building a prediction model on the performance score for each target, the zero-order correlations were examined. It turned out that the similarity in age did not have a meaningful correlation with the performance score (*r* between -.104 and .079) so that this variable was excluded from the regression analysis. Assumed similarity did also not correlate substantially with any performance measure. However, it was retained in the analysis because Cronbach's model also considered a possibly meaningful interaction effect of assumed and real similarity. The regression analysis included both continuous and dichotomous variables, the latter were already coded as dummy variables. This needed to be accounted for when interpreting the resulting regression coefficients. Table 8.28 presents the results of the regression analysis. The analysis showed only some small effects on the performance in the scenarios. Only the performance in the scenario Renate was significantly predicted by the similarity indicators with about 8.1 % of explained variance (R^2_{adj}). The further prediction models did not reach significance, the amount of explained variance (R^2_{adj}) ranged between 0 % and 2.5 %.

With regard to the single predictors, the beta-coefficients did not show a coherent pattern of prediction and varied substantially between the scenarios. Many zero-order correlations and beta-coefficients were negative in sign. The expected interaction effect between the assumed similarity and the real similarity indicators could only be detected in the scenario Christoph (CP). Here, assumed similarity showed a suppressing effect enhancing the prediction of the similarity in the personality profiles.

Table 8.28

Multiple Regression Analysis to Predict the Performance in the General Target-Related Social Understanding Scores (Target Scoring) by Similarity

CR	**1 (RF)**		**2 (BS)**		**3 (CK)**		**4 (CP)**		**5 (KL)**		**6 (FB)**		**7 (HR)**		**8 (MM)**	
PR	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**
AssSim	-.018	-.014	-.044	-.027	-.095	-.082	-.050	-.085	.068	.068	-.006	-.007	-.037	-.028	.023	.033
Gender	.159	.189*	-.098	-.129	-.059	-.062	.059	.028	-.047	-.047	-.024	-.040	-.076	-.057	.153	.119
Parent-hood	.009	-.027	-.098	-.104	.000	-.054	.035	.014	.071	.071	-.086	-.064	-.066	-.037	.174	.118
Educa-tion	.245	.243**	-.015	.005	-.070	-.085	-.056	-.063	.180	.180*	-.112	-.094	-.043	-.029	.129	.097
Perso-nality	-.121	-.138	-.138	-.119	.067	.078	.112	.135	.026	.026	-.065	-.046	.073	.062	.011	.001
R^2 (R^2_{adj})	**.106 (.081)**		**.040 (.013)**		**.023 (.000)**		**.025 (.000)**		**.046 (.018)**		**.020 (.000)**		**.014 (.000)**		**.052 (.025)**	
F	4.161**		1.465		.818		.907		1.676		.716		.489		1.937	
DF*** (reg/res)	5/175 (181)		5/174 (180)		5/176 (182)		5/176 (182)		5/175 (181)		5/175 (181)		5/176 (182)		5/176 (182)	
P	.001		.204		.540		.478		.143		.612		.785		.090	
Δ R^2 sympathy (β)	**.001**		**.125 (-.407**)**		**.024 (-.202*)**		**.002**		**.007**		**.073 (-.311**)**		**.008**		**.001**	

Note. * p < .05, **p < .01, *** N in parentheses
PR = predictor, CR = criterion (general target-related performance score), AssSim = assumed similarity, predictors are all indicators of similarity (higher value indicating higher similarity)

Interestingly, adding *perceived sympathy* into the regression model enhanced the prediction in three scenarios. The perceived sympathy was assessed by only one item per scenario on a 7-point rating scale from 1 (*not at all sympathetic*) to 7 (*very sympathetic*). The last line in Table 8.28 presents the increase in R^2 and the beta coefficients indicating the significance of the single predictor. The zero-order correlations of 'perceived sympathy' with the general performance in the scenarios for BS, CK, and FB were r = -.313 / -.187 / -.235 respectively, indicating a negative effect of 'perceived sympathy' on the performance in these three scenarios. In general, the amount of explained variance was still low. However, some interesting effects could be discovered when only specific scenarios were regarded separately supporting the importance of carefully considering the selection of targets in terms of heterogeneity and number.

Effect of Similarity on Social Understanding Scales Scored by Correlations-Based Score

Snodgrass (2001) claimed that correlations-based scoring methods compensate for the differential use of rating scales by subjects and targets. At the same time, it was argued that this type of scoring is influenced by the use of guessing strategies in social judgments (i.e., stereotypes as the foundation for heuristic judgments). Thus, it should be explored whether correlations-based scores were predicted by the assumed and the real similarity between subjects and targets to a larger extent than the conventional target scoring procedure. Table 8.29 presents the results of the same multiple regression analysis as presented in Table 8.28, this time applying the correlation-based scores as the criterion variables. In order to allow a direct comparison with the aforementioned analysis, the same predictors were applied. The fact that the reliabilities of these scales were very low (see preceding Chapter) must be taken into account. This resulted in a limited capability to predict systematic variance.

Except for two scenarios, the regression model predicted a significant amount of variance in the social understanding tasks with R^2_{adj} between .075 and .102. Again, the beta coefficients did not show a systematic effect on performance. No variable showed a significant contribution across all scenarios. The assumed similarity showed a negative effect on the performance in the scenarios Conny (CK) and Christoph (CP). Contrarily, it showed a positive effect on performance in the scenario Friedrich (FB).

In general, larger amounts of variance were predicted in these scores by the indicators of assumed and real similarity. Adding ‘perceived sympathy’ to the model, the prediction was significantly enhanced for three scenarios (i.e., CK, CP, and FB), with $\Delta R^2 = 0.148$, / 0.029 / 0.026 respectively. Corresponding to the other predictors, sympathy effected the performance in different directions for different scenarios. The beta coefficients (see Table 8.29) for the scenarios CK and CP for ‘perceived sympathy’ were negative in sign and the zero-order correlations showed the same direction ($r_{sv.symp}$ = -.436 / -.236; respectively for CK and CP). Performance in the scenario Friedrich (FB), on the other hand, was positively affected by ‘perceived sympathy’ with a positive zero-order correlation ($r_{sv.symp}$ = .240).

Table 8.29

Multiple Regression Analysis to Predict the Performance in the General Target-Related Social Understanding Scores (Correlations-Based Scoring) by Similarity

CR	1 (RF)		2 (BS)		3 (CK)		4 (CP)		5 (KL)		6 (FB)		7 (HR)		8 (MM)	
PR	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**	**r**	**β**
AssSim	.183	.141	.023	.012	-.171	-.202**	-.203	-.200*	.056	-.004	.178	.205**	.083	.051	.021	.029
Gender	.200	.144	-.143	-.148*	.035	-.007	.090	.081	.227	.264*	-.033	-.123	.199	.195*	-.124	-.083
Parent-hood	-.138	-.112	.028	-.076	-.245	-.277**	.043	-.015	.004	.094	-.167	-.173*	.065	-.003	-.149	-.151*
Educa-tion	.054	.095	.159	.151*	.047	-.004	-.013	-.020	.061	.049	-.191	-.149*	.171	.168*	.017	.068
Perso-nality	.098	.031	.213	.21**	.014	.036	-.049	-.006	.163	.201*	.010	.077	.043	.073	-.031	-.039
R^2 (R^2_{adj})	**.078 (.052)**		**.088 (.062)**		**.102 (.077)**		**.048 (.021)**		**.096 (.070)**		**.099 (.073)**		**.075 (.048)**		**.036 (.008)**	
F	2.973*		3.352**		4.018**		1.765		3.710**		3.827**		2.843*		1.307	
DF*** (reg/res)	5/175 (181)		5/174 (180)		5/176 (182)		5/176 (182)		5/175 (181)		5/175 (181)		5/176 (182)		5/176 (182)	
P	.013		.006		.002		.122		.003		.003		.017		.263	
Δ R^2 symp (β)	**.019**		**.000**		**.148 (-.496**)**		**.029 (-.221*)**		**.004**		**.026 (.190*)**		**.000**		**.008**	

Note. * $p < .05$, ** $p < .01$, *** N in parentheses
PR = predictor, CR = criterion (general target-related performance score),
AssSim = assumed similarity, symp = perceived sympathy, predictors are all indicators of similarity (higher value indicating higher similarity)

To summarize, no general conclusion about the effect of similarity or sympathy on the performance in the social understanding tasks could be undertaken. Many diverse effects occurred in the analysis showing no systematic pattern of prediction by single predictor variables. This finding added to the results from the analysis of gender differences related to different target genders on the performances. Both findings point towards the importance of applying many and heterogeneous targets in order to balance effects of single target persons within the performance scales.

However, it still needs to be considered that different items were underlying the different scenarios and the comparability of the scales was restricted. Nevertheless, the general scenario scores were highly correlated (see above). They contained only items that were included in the final content-related scales which had shown good internal consistencies.

Thus, there was at least some evidence that the performances within the scenarios was not only determined by different items but also by the same underlying ability domain.

If this assumption was accepted, it could be reasoned in accordance with Snodgrass (2001) that the use of heuristics play a more prominent role when correlations-based scores are applied. However, these heuristics contributed diversely to the performance in the social understanding tasks, frequently resulting in lower scores.

8.4.4.5 Exploring the Faceted Structure of Social Understanding – Research Question 3E

The social understanding tasks were based on a 4 x 3 x 2 design cross-classifying four content domains (i.e., written and spoken language, pictures, and videos; V, A, P, and F, respectively), three modality domains (i.e., emotions, cognition, and relationships; E, C, and R, respectively), and two setting domains (i.e., private and public; PR and PU, respectively). This design was realized through the application of eight scenarios, each related to one target person. Originally, the design served as a methodological tool to balance possibly relevant variance components, to enhance representativeness of task material, and to control for method-related variance. However, literature suggested that some of the classificatory elements could as well represent meaningful ability domains that share systematic common variance. For example, academic intelligence theories distinguish between content ability domains related to verbal, figural-spatial or numerical contents (Carroll, 1993; Jäger, 1982). Tests of emotional abilities focus only on emotional task contents purporting that these represent a meaningful ability domain. However, tests of emotional intelligence conventionally do not systematically vary other modalities. Most of the tests of social abilities also do not systematically vary the contents (i.e., most of them rely only on written language) or the settings (see Table 5.3 and Chapter 5.3).

The tasks applied in the present study allowed the investigation of the social understanding structure by relying on the three potential ability facets of contents, modalities, and settings. Whether the data supported a general social understanding factor beyond the specific ability factors, was also investigated.

As a first step, the 4 x 3 x 2 design was realized by building the respective 24 cells. The item-total correlations and the reliabilities of the scales were examined. Table 8.30 presents the results from this analysis showing the number of items in the cells, the descriptives, the empirically determined reliability coefficients, and the estimated reliabilities. The reliability estimation was conducted by applying the Spearman-Brown-Formula. The applied test length was that of the content-related scales of social understanding with an average item number of 60.

Table 8.30

Psychometric Properties of the Social Understanding Cells Cross-Classifying Contents, Modalities, and Settings

Cell*	α_{emp}	Item Count	% Items selected	$r_{tt\ est}$	M	SD
SU_VEPR	.400	14	54	0.74	-1.991	.498
SU_VCPR	.434	18	90	0.72	-2.463	.485
SU_VRPR	.293	6	46	0.81	-2.220	.623
SU_AEPR	.525	22	65	0.75	-2.242	.469
SU_ACPR	.483	18	75	0.76	-2.353	.474
SU_ARPR	.533	20	87	0.77	-1.816	.429
SU_PEPR	.477	16	70	0.77	-2.698	.540
SU_PCPR	.300	9	56	0.74	-2.398	.671
SU_PRPR	.463	15	83	0.78	-2.212	.499
SU_FEPR	.365	9	53	0.79	-1.991	.540
SU_FCPR	.451	13	72	0.79	-2.217	.553
SU_FRPR	.529	17	85	0.80	-1.990	.442
SU_VEPU	.630	20	71	0.84	-2.815	.506
SU_VCPU	.395	13	65	0.75	-2.846	.543
SU_VRPU	.314	5	63	0.85	-1.869	.637
SU_AEPU	.619	23	77	0.81	-2.206	.485
SU_ACPU	.334	9	38	0.77	-2.631	.645
SU_ARPU	.366	10	59	0.78	-1.893	.482
SU_PEPU	.396	18	72	0.69	-1.984	.443
SU_PCPU	.458	16	80	0.76	-2.198	.493
SU_PRPU	.358	10	91	0.77	-1.776	.444
SU_FEPU	.581	25	81	0.77	-2.109	.447
SU_FCPU	.571	13	81	0.86	-2.731	.645
SU_FRPU	.501	16	80	0.79	-1.680	.474

Note. * first letter: content domain, second letter: modality domain, third and forth letter: setting domain, SU = social understanding, V = written language, A = spoken language, P = picture, F =video, E =emotion, C = cognition, R = relationships, PR = private, PU = public, α_{emp} = empirical α, $r_{tt\ est}$ = estimated r_{tt}

Prior to the reliability estimation, the item-total correlations were inspected. Items with a negative value were excluded from the analysis. Table 8.30 presents the percentage of items that were included in the final cells. It is worth noting that the percentage of items selected was generally larger than that of the overall content-related scales (see Chapter 8.4.2.1). Between 41 and 44 % of the items in the scales SUv, SUa, and SUp showed positive item-total correlations. Only in the SUf scale, 81 % of the items correlated positively with the total scale (see Table 8.9). Moreover, the empirically determined reliabilities were already high, accounting for the reduced number of items per cell. Consequently, the cells showed reasonable estimated reliability coefficients of $r_{tt\ \mathrm{est}}$ between .69 and .86. The zero-order correlations were inspected and did not show a consistent pattern. Most of the correlations were positive in sign ranging from zero to .481. Only some were negative. The mean correlations between the cells was low with $r = .14$.

Prior to investigating the faceted structure in combined models with all facets, the structure within one facet needed to be examined. Therefore, confirmatory factor analysis was applied to test a series of several models. The rational of analysis is presented in Table 8.31.

All factors of one facet were postulated to be correlated. First, a general factor model with loadings of all cells was analyzed (Model U). Models V and W investigated the structure of the content facet. Model V postulated four ability factors according to the design. Model W established a two-factor structure including a language-based and a language-free factor (Model W) as identified in the faceted structure model of social intelligence (see Model I in Figure 8.11). Models X examined the factor structure of the modality facet postulating a three-factor solution with all three modalities (Model X). Finally, Model Y established a two-factor model of the two setting domains. Table 8.31 presents the rationale for analysis and the summary of fit statistics from confirmatory factor analysis.

None of the models showed a good data fit which was not surprising due to the low correlations between the cells. The loadings of the cells on the various latent factors were mostly positive in sign with the exceptions of two of the 24 cells which had slightly negative loadings of between -.03 and -.07 in all models (i.e., the SU_VRPU and SU_ARPU). The remaining loadings on all latent factors ranged between .22 and .73.

Table 8.31

Fit Statistics for Confirmatory Factor Analyses of the Structure of the Social Understanding Facets

Model	χ^2	DF	p (χ^2)	CFI	RMSEA	SRMR	CI RMSEA*
U: General factor model of SU	532.348	252	<.001	.624	.078	.091	[.069; .087]
V: 4-Factor model of content facet (V-, A-, P-, and F-factors)	524.908	246	<.001	.626	.079	.091	[.070; .088]
W: 2-Factor model of content facet (lb- and lf-factor)	531.469	251	<.001	.624	.079	.091	[.069; .088]
W1: 2-Factor model of content facet (V- and A/P/F-factor)	528.033	251	<.001	.629	.078	.091	[.069; .087]
X: 3-Factor model of modality facet (E-, C-, and R-factor)	501.124	249	<.001	.662	.075	.091	[.065; .084]
X1: 2-Factor model of modality facet (E/C- and R-factor)	509.019	251	<.001	.654	.075	.091	[.066; .084]
Y: 2-Factor model of setting facet (PR- and PU-factor)	532.349	251	<.001	.643	.079	.091	[.069; .088]

Note. N = 182, * CI = 90%
SU = social understanding, v = written language, a = spoken language, p = picture, f = video, lb = language-based, lf = language-free factor, e =emotion, c = cognition, r = relationships, pr = private, pu = public

However, some conclusions were possible about the adequate number of factors per facet. Model V investigated the structure of the content facet and showed a perfect correlation of r = 1.00 between the pictorial and the video-based factor and a very high correlation of r = .90 between the pictorial and the auditory factor. The written language factor correlated with the other factors with r = .76 / .77 / .85 (with A, P, and F, respectively). From these high correlation indices computed, it was not clear whether the factors were structured as postulated by the original design. Moreover, data fit was not significantly better than that of the general factor model (Model V; χ^2-difference = 7.440, df = 6, n.s.).

Two more models were hypothesized representing first a two-factor solution (Model W) with a language-based and a language-free factor comparable to the faceted social intelligence structure from Model I (see Figure 8.11). Due to the slightly lower correlations of the written language factor compared to the other correlations, Model X1 postulated a two-factor solution with a separate written language factor and a combined content factor with loadings of all cells belonging to the spoken language, pictorial, and video-based content domains. The latent factor intercorrelations were still very high. The language-based and language-free factors in Model W correlated with r = .95, the written language factor with the factor of the remaining contents $r = .86$. The χ^2-differences test still showed that Model U fitted the data better than Model W. However, the data fit of Model W1 was significantly better than that of Model U (χ^2-difference = 4.315, df = 1, $p < .05$) suggesting the separability of the verbal content factor.

Models X examined the factor structure of the modality facet. The three-factor solution in Model X showed a perfect correlation of $r = 1.00$ between the emotion's and the cognition's modality domains. The correlations of the modality relationships with the emotions and cognitions domain were $r = .60$ and .83, respectively. Therefore, Model X1 was established postulating a two-factor solution with a combined emotion and cognition modality factor and a relationships factor. The two factor correlated with $r = .72$. Finally, Model Y examined the settings facet and did not find evidence for separable ability domains related to the private and public task settings. Both factors correlated perfectly with $r = 1.00$.

So far, the structure of only one facet at a time was examined. Now, the combined facetted models of social understanding should be established. The settings facet will be omitted from the analysis because data showed that no differentiation was possible between the two setting domains. Thus, a reduced faceted design was applied relying on cells on a higher level of analysis. The design relied on the 4 x 3 classification of four contents and three modalities. Another reliability analysis was conducted with these 12 cells. The results are presented in Table 8.32 including the number of items in the cells, the descriptives, the empirically determined reliability coefficients , and the estimated reliabilities. The reliabilities were estimated by applying the Spearman-Brown-Formula to a test length of 60 items.

Again, the reliabilities were relatively high for the small item number in the cells. Moreover, the number of items that were omitted due to a negative item-total correlations was low compared to the overall content-related scales (see Chapter 8.4.2.1). In general, both reliability analyses based on the cells suggested that the scales were more internally consistent and items were more homogeneous when the entity of analysis was smaller.

Table 8.32

Psychometric Properties of the Social Understanding Cells Cross-Classifying Contents and Modalities

Cell*	α_{emp}	Item Count	% Items selected	$r_{tt\ est}$	M	SD
SU_VE	.638	34	63	0.76	-2.504	.397
SU_VC	.583	29	71	0.74	-2.622	.430
SU_VR	.261	14	67	0.60	-1.965	.384
SU_AE	.718	45	70	0.77	-2.187	.418
SU_AC	.471	33	69	0.62	-2.515	.363
SU_AR	.577	32	80	0.72	-1.821	.348
SU_PE	.572	32	67	0.71	-2.367	.404
SU_PC	.515	26	72	0.71	-2.271	.439
SU_PR	.483	25	86	0.69	-1.958	.355
SU_FE	.646	36	75	0.75	-2.103	.396
SU_FC	.623	25	74	0.80	-2.517	.490
SU_FR	.636	33	83	0.76	-1.839	.371

Note. * first letter: content domain; second letter: modality domain, SU = social understanding, V = written language, A = spoken language, P = picture, F =video, E = emotion, C = cognition, R = relationships
α_{emp} = empirical α, $r_{tt\ est}$ = estimated r_{tt},

The zero-order correlations between the cells are presented in Table 8.33. Only the correlations below the diagonal are relevant at present. The correlations between the cells were consistently high with only very few zero or slightly negative correlations. The cells SU_VE, SU_PR, and SU_FR showed some systematically lower correlations with other cells. The correlations were generally larger than those between the previous cells which additionally varied the setting (mean *r* = .26 compared to .14 in the preceding analysis).

Table 8.33

Correlations between the Social Understandings Cells Cross-Classifying Contents and Modalities

Cell[a]	SU_VE	SU_VC	SU_VR	SU_AE	SU_AC	SU_AR	SU_PE	SU_PC	SU_PR	SU_FE	SU_FC	SU_FR
SU_VE		.376**	.009	.186*	.131	.269**	.340**	.146*	.300**	.363**	.448**	.303**
SU_VC	.399**		.219**	.450**	.334**	.484**	.441**	.325**	.389**	.612**	.503**	.396**
SU_VR	.021	.200**		.250**	.228**	.202**	.176*	.213**	.220**	.261**	.252**	.272**
SU_AE	-.100	.291**	.145		.331**	.248**	.286**	.300**	.232**	.502**	.413**	.329**
SU_AC	.089	.345**	.099	.376**		.108	.357**	.361**	.098	.349**	.394**	.217**
SU_AR	.161*	.380**	.247**	.217**	.004		.248**	.214**	.395**	.275**	.366**	.518**
SU_PE	.156*	.336**	.167*	.324**	.297**	.263**		.096	.151*	.357**	.478**	.321**
SU_PC	-.009	.184*	.189*	.364**	.283**	.253**	.155*		.245**	.372**	.346**	.289**
SU_PR	.267**	.312**	.269**	.072	-.037	.419**	.167*	.208**		.335**	.378**	.381**
SU_FE	.165*	.514**	.161*	.522**	.377**	.224**	.328**	.374**	.200**		.514**	.336**
SU_FC	.431**	.458**	.221**	.238**	.317**	.298**	.355**	.295**	.308**	.454**		.448**
SU_FR	.275**	.310**	.288**	.163*	.065	.504**	.280**	.272**	.395**	.241**	.442**	

Note. N = 182; * $p < .05$; ** $p < .01$
[a] first letter: content domain; second letter: modality domain
correlations based on cell residuals (with target variance partialled out) above diagonal
v = written language, a = spoken language, p = picture, f =video, e =emotion,
c = cognition, r = relationships

Eventually, confirmatory factor analysis should be used to examine the faceted structure of social understanding as operationalized by the 12 cells. The rationale of the analysis and the fit statistics of the models are presented in Table 8.34. First, a general factor model was established (Model Z) with loadings of all of the cell indicators on one general factor. The second model postulated the entire 4 x 3 design with four correlated content factors and three correlated modality factors (Model AA). According to the preceding results (see Model V), further models combined the emotions and the cognitions factor on the modality facet. Consequently, Model AB relied on a 2 x 2 design with two correlated modality factors (EC and R) and two correlated content factors (i.e., a V-factor and a combined factor of A-, P-, and F-contents). Model AC postulated a language-based and a language-free factor on the content facet and two modality factors. All factors were correlated. Since the content factors showed rather high intercorrelations in the preceding analysis, the next model established one general factor with the loadings of all of the indicators and two uncorrelated modality factors (hierarchical solution; Model AD). This model tested whether a differentiation into different content factors was necessary. A final model in the course of this analysis relied on the 2 x 2 faceted structure as identified in Model AC. Additionally, a higher order general social understanding factor with the loadings of all of the indicators was introduced. The correlations between the ability factors were omitted (Model AE). This model tested whether data supported a hierarchical faceted structure of social understanding.

Table 8.34

Fit Statistics for Confirmatory Factor Analyses of the Faceted Structure of Social Understanding

Model	χ^2	DF	p (χ^2)	CFI	RMSEA	SRMR	CI RMSEA*
Z: General factor model of SU	202.708	54	<.001	.722	.123	.096	[.105; .141]
AA: 4x3-Facetted model (4 contents v, a, p, f and 3 modalities e, c, r)	54.101	33	.012	.961	.059	.047	[.028; .087]
AB: 2x2-Facetted model (v- and a/p/f-content factors, and 2 modality factors e/c and r)[a]	65.422	41	.009	.954	.057	.050	[.029; .082]
AC: 2x2-Facetted model (2 contents lb and lf and 2 modalities e/c and r)	62.491	40	.013	.958	.056	.051	[.026; .081]
AD: Hierarchical model (general content factor and 2 modalities e/c and r)	69.821	42	.004	.948	.060	.054	[.034; .085]
AE: Hierarchical model with 2x2-facetted design (uncorrelated factors) and general SU-factor	52.149	31	.010	.961	.061	.047	[.030; .089]
AF: 2x2-Facetted model AC based on cell residuals (target variance controlled)[a]	58.866	41	.035	.971	.049	.041	[.014; .075]

Note. N = 182; * CI = 90%; [a] two error terms constrained to be equal
SU = social understanding, v = written language, a = spoken language, p = picture, f = video, lb = language-based, lf = language-free factor; e =emotion, c = cognition, r = relationships, pr = private, pu = public

Data fit for the general factor model was poor (Model Z; CFI = .722; χ^2 = 202.708; p < .001). The loadings were all positive and varied between .34 and .69. Model AA postulated the original 4 x 3 design and showed substantially better data fit (CFI = .961; χ^2 = 54.101; p=.012). The loadings on the modality factors were rather homogeneous and ranged between .38 and .71. Again, the emotion and cognition factors correlated perfectly with *r* = 1.00 and were combined in further analysis. The relationships factor correlated with the emotions and cognitions factors with r = .58 and .66 respectively. The loadings on the content-related factors, however, were not consistent; a few loadings were slightly negative. The loadings on the written and spoken language factor ranged between -.08 and .93, and those on the pictorial and video-based content factors ranged between -.23 and .34. Thus, the factors could not be identified as coherent ability factors of the respective content domains.

According to the preceding results, the next model postulated a 2 x 2 faceted structure with a separate written language content factor and a combined factor of the remaining contents. Two modality factors were established (i.e., E/C and R), all factors belonging to one facet were correlated. The model showed reasonable data fit close to the fit of the 4 x 3 faceted structure model. One error term appeared at the lower bound so that an equality constraint was introduced. Models AA and AB were nested. The χ^2-differences test showed no significant better fit for Model AA (χ^2-difference = 11.321, df = 8, n.s.) and the more parsimonious Model AB was accepted supporting the combination of the ability factors E and C and the content factors A, P, and F. The correlation between the content factors, however, was again very large (*r* = .92). To explore the structure more closely, another faceted model was established (Model AC). This model postulated a different structure on the content facet and combined the written and spoken language factors into a language-based content factor. The pictorial and video based factors were combined into a language-free content factor. Again, the emotions and cognitions modality factors were combined. The data fit was slightly better than that for Model AB (CFI = .958; χ^2 = 62.491; p=.013). Moreover, the relation of χ^2 to the degrees of freedom was slightly better than that of Model AB. Again, the χ^2-difference test showed no significant better fit for Model AA (χ^2-difference = 8.390, df = 7, n.s.) so that the more parsimonious Model AC was supported. The content factors still showed a rather large intercorrelation (*r* = .84). Because of the necessary equality constrained in Model AB and the slightly worse relation of χ^2 to the degrees of freedom compared to Model AC, Model AC was accepted as the best model to utilize.

Figure 8.18 presents the standardized solution showing the loadings and factor intercorrelations. The modality factors exhibited a reasonably large correlation (*r* = -.48). The negative sign corresponded with the loadings on the two factors which were negative on the emotions/ cognitions factor and positive on the relationships factor.

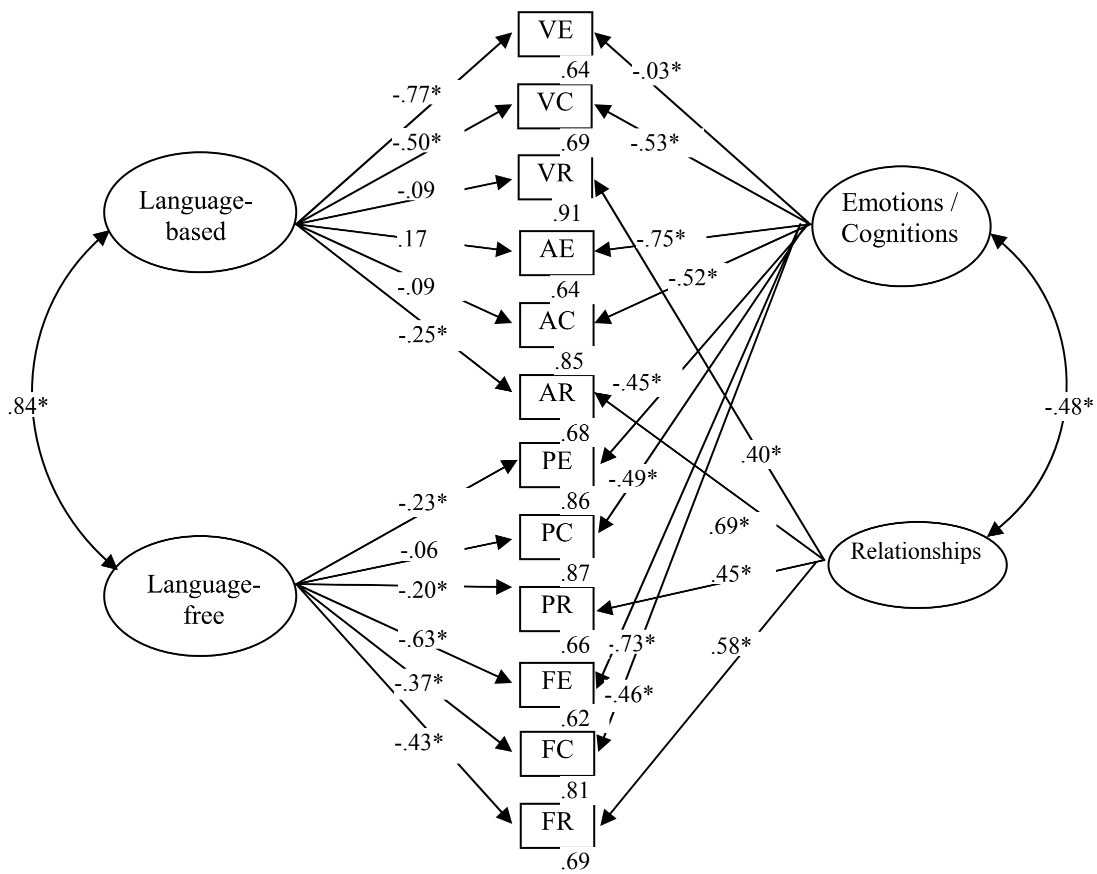


Figure 8.18

Standardized Solution of Faceted Model of Social Understanding (Model AC)

Note. CFI = .958; χ^2 = 62.491, p=.013; * p < .05; error terms are enclosed with the manifest variables
v = written language contents, a = spoken language contents, p = pictorial contents, f = video-based contents, e = modality emotion, c = modality cognition, r = modality relationships

The loadings particularly on the content factors, however, were very heterogeneous. Except for one indicator (i.e., .17; the spoken language contents emotion modality), all loadings were of equal direction but varied largely between -.06 and -.77. The loadings on the modality factors were more homogeneous showing only one outlying loading (-.03 for cell VE). The model clearly supported the differentiation into the modality factors and their coherence as an ability domains. The factor intercorrelations between the content factors, however, raised doubts concerning the necessity of this differentiation. Moreover, the loading pattern did not support the content factors as meaningful and useful content domains.

The next model tested whether the content facet in fact represented a necessary differentiation. Therefore, a model was established with two modality factors and one general factor consisting of the loadings from all of the cell indicators (Model AD). The modality factors were kept uncorrelated. The model showed reasonable data fit (CFI = .948; χ^2 = 69.821; p=.004). A χ^2-differences test showed a significantly better fit for Model AC (χ^2-difference = 7.330; df = 2, p < .05) supporting the differentiation into the two content domains.

Another hierarchical model was postulated relying on the faceted 2 x 2 structure supported in the previous analysis (Model AC). Additionally, one general social understanding factor with loadings of all of the indicators was introduced where the ability factors of the two facets were kept uncorrelated. The model also showed reasonable data fit (CFI = .961; χ^2 = 52.149; p=.010). The loadings on the general social understanding factor were rather homogeneous and ranged between .23 and .73. The loadings on the modality factors remained stable compared to Model AC. The loadings, however, on the content-related factors changed meaningfully (i.e., loadings on the language-based content factor were between -.49 and .25; loadings on the language-free content factor ranged between -.36 and -.08). The loadings did not exhibit a consistent pattern, particularly on the language-based contents; thus this factor could not be identified as a coherent ability domain.

In summary, the models supported a modality facet that differentiated between relationships and emotions / contents as meaningful ability factors. The loadings on the modality factors were generally consistent. Contrarily, evidence for a meaningful content related ability facet was weaker. Particularly, the loadings on the language-based content factors were very heterogeneous, making this factor unable to be interpreted. However, a general social understanding factor was supported by the consistently large loadings of all of the indicators.

Effect of Target-Related Variance – Research Question 3E2

The social understanding tasks (i.e., the scenarios) were related to eight different target persons. The large correlations between all of the scales or between the previously identified ability factors might have been elevated by the common variance related to information from the scenarios (i.e., information about the same target persons). Thus, the effect of the target related variance on the social understanding structure was to be explored in this section.

Originally, it was intended to control for the common target variance by establishing uncorrelated target related factors in addition to the previously established faceted structure. Therefore, cell indicators were built relying only on the cross-classification of contents, modalities, and targets. The same principle for the cell building was applied as in the preceding analysis. The estimated reliabilities of the cell indicators ranged between .44 and .92 (based on 5 – 23 items). The cell intercorrelations within the scenarios did not suggest underlying common variance in most of the scenarios. Only the scenarios Renate (RF) and Hannah (HR) showed meaningful within-scenario correlations of *r* between .126 and .406. The correlations for the remaining scenarios ranged from *r* = -.201 to *r* = .348.

Attempts to conduct confirmatory factor analysis relying on these cells was not successful and most of the models did not converge. Therefore, the model should be approximated by exploratory factor analysis in order to extract factor scores of the target related factors. The procedure applied was as follows. First, the factor score of the first principle component of all of the indicators was saved. Thus, the common variance of all of the indicators was determined which could be labeled as a general social understanding factor. This factor score was partialled out of every single cell indicator. Through this step, the desired common social understanding variance was controlled for. This step was intended to correspond with the establishment of a general factor of social understanding besides the eight target related factors. The latter would only assemble variance purportedly unrelated to the common social understanding variance. Without this step, the target related factors would also include general social understanding variance which was not the desired outcome. Subsequently, in order to build indicators of the common target variance per scenario, eight further exploratory factor analyses were conducted and the factor scores were saved. Each one of the factor analyses entered only the residuals of all of the indicators related to one scenario. These factor scores were sought to contain only the variance common to one target person (i.e., without any common social understanding variance) only. Finally, these factor scores were partialled out of the previously built cells (i.e., cross-classifying contents and modalities) in order to examine the faceted structure without the influence of the target related variance.

The intercorrelations between the cell residuals are presented in Table 8.33 above the diagonal. All correlations increased in size compared to those between the original cell indicators. The two-faceted model with 2 x 2 factors of contents and modalities (Model AC) was replicated relying on the cell residuals (see Model AF in Table 8.34). The model showed good data fit (CFI = .971; χ^2 = 58.866; p=.035). However, one error term was at the lower bound so that an equality constraint on two error terms was introduced. The loadings on the modality factors were stable compared to Model AC. The loadings on the emotions / cognitions factor ranged between .48 and .76, and on the relationships factor between .41 and .72. The factor intercorrelation was higher than in Model AC (*r* = .74). The loadings on both content factors, however, revealed a heterogeneous pattern showing both positive and negative loadings ranging from -.48 to .38; an interpretation of the two factors was not possible. The factor intercorrelation also changed substantially from .84 in Model AC to .29 in Model AF. This was surprising because the underlying cell intercorrelations were also larger between the cells of the same contents.

In summary, the results suggested that controlling for common target variance seemed to replicate and support the modality facet with two factors. The content related factors were not supported by data, which cannot be explained at this point. However, the way the factor scores of the common target variance were composed was certainly not the most ideal way. It would have been better to extract the factor scores from confirmatory factor analysis since this would have excluded the error variance from the factor scores which is now still included. Thus, it did not seem surprising that the residuals of the cell indicators intercorrelated to a large extent because the unsystematic variance was removed by partialling out the factor scores of the target factors.

Interaction effect of tasks contents and modalities on performance in social understanding tasks – Research Question 3E3

The third research question related to the facetted design of the social understanding tasks, referred to the question of whether performance varied between the cells. Archer and Akert (1980) investigated the effect of task material on the performance of social understanding tasks in a between-subjects design. Each group of subjects worked on the same tasks but was provided with different task material (i.e., different cues). One group was provided with the full material (i.e., including audio and pictorial information). One group was provided only with video information, one with auditory information. Results showed that there was no meaningful general influence of the type of cues or the task material on the accuracy. However, some specific effects were discovered showing that some task material favored performance in some aspects. In certain cases, subjects'performance reached a level comparable to the "full material condition" when only one part of a scene or just one piece of information was provided.

The design of the social understanding tasks in the present study did not allow for the replication of this analysis. A difference between the mean performance in the cells of the social understanding tasks could not be attributed unambiguously to an interaction of task material with the queried modality but could also be an effect of different tasks underlying each cell. However, the items within the cells were also heterogeneous (see Table 6.3 in

Chapter 6.2.2.1) with the only common element, for example, being the judgment of emotions based on pictorial material. In any event, the investigation of the mean differences between the cells still appeared interesting and could turn out to be relevant for possible subsequent test modifications.

Thus, the mean differences between the cells of the 4 x 3 design were investigated by applying a two-factor analysis of variance involving two repeated measurement factors (i.e., one factor with four contents and one factor with three modalities). The means and standard deviations of the cells are included in Table 8.32. Table 8.35 presents the results of the ANOVA with two repeated measurements factors. Both repeated measurement factors and the interaction effect reached significance. Figure 8.19 shows the means of the cells. Post hoc tests revealed significant mean differences between most of the paired comparisons except for the comparisons between *(a)* the cells SU_VR and SU_PR, *(b)* the cells SU_AR and SU_FR, and *(c)* the cells SU_AC and SU_FC.

Table 8.35

Analysis of Variance with Two Repeated Measurement Factors (Contents and Modalities)

Source of variance	Sums of square	DF	Mean squares	F	Probability	Eta²
Repeated measures (contents)	15.083	3	5.028	43.870**	<.001	.028
Repeated measures (modalities)	129.689	2	64.844	445.753**	<.001	.237
Interaction effect	17.890	6	2.982	26.414**	<.001	.033

Note. N = 182

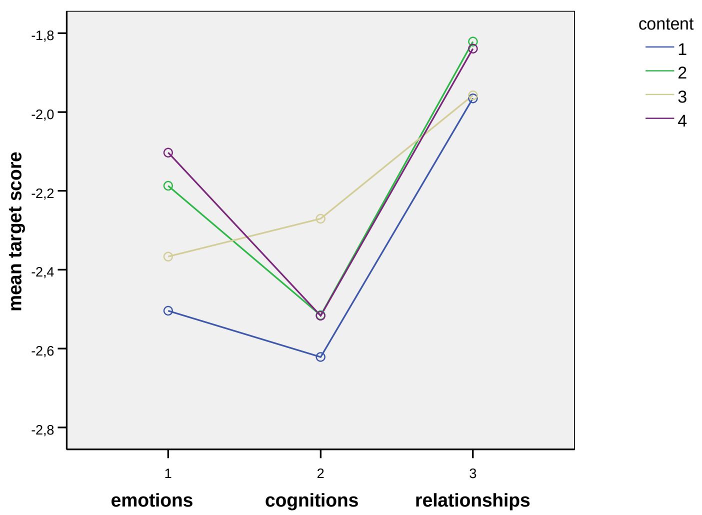


Figure 8.19

Mean Performance in the Cells of the 4 x 3 Facetted Design of Social Understanding Tasks

Note. 1 = written language, 2 = spoken language, 3 = pictures, 4 = videos

In general, there was a large effect of the modality facet, which exhibited better performance for the judgment of relationships and lower performance for the judgment of cognitions. The effect of the content facet was small but significant and showed the lowest performance in cells based on written language. This finding replicated the results from the analysis of the descriptives of the social understanding scales (see Chapter 8.4.2.1 and Table 8.11).

The interaction effect in the surprisingly high mean performance of the cell SU_PC clearly suggested an effect of the task construction. The pictorial cells showed the second lowest performance in judging emotions and relationships but the best performance in judging cognitions. There was no plausible explanation for this effect except for the fact that the items for this cell were possibly constructed to be easier because of the assumption that it could be harder to judge emotions based only on pictorial information.

8.4.4.6 Process During the Tasks of Social Understanding – Research Question 3F

This last section regards the processes that underlie the social understanding tasks. Some researchers distinguished between the role of perception, interpretation, and knowledge in social judgments (Bless et al., 2004; Buck, 1983) and some were concerned with the necessary amount of extrapolation from the task material to the requested output (i.e., direct vs. mediated perception; see Buck, 1983; Gage & Cronbach, 1955). For a more detailed description of these processes see Chapter 4.3.3.1 and 5.2.4.1. The definitions underlying the test construction clearly suggested that only the interpretation of the task material should determine the performance in the social understanding tasks. Knowledge requirements should only be included to a minimum extent. Consequently, both perception (i.e., of the relevant cues in the task material) and interpretation requirements should be addressed by the social understanding tasks. Thus far, it was not clear whether these postulated processes were in fact realized by the tasks. The tasks provided substantial background information which could be used to assemble a knowledge base about the target person. The consecutive questions are whether test takers only rely on information from the task material for judging the mental states of target persons or how much knowledge about target persons or the presented situations contributed to their performance?

A study which would be capable to provide answers to these questions would require a different design than the one of the present study. How such designs could be established will again be addressed in Chapter 10. At present, these questions were approached in two different ways. First, the present analysis examined whether there was an increase in the accuracy of judgments during the course of the scenario tasks. If an effect of the course of the tasks within one scenario on the accuracy could be observed, this could then be attributed to the assembling of an underlying knowledge structure (Research Question 3F1). Second, the relationship of performance in the social understanding tasks to the long term memory (LTM) task was explored (Research Question 3F1).

Performance During the Course of the Scenarios – Research Question 3F1

As a first step, the correlations of the mean performance for each item with the item position in terms of the underlying scene within the course of the scenarios was assessed. Table 8.36 presents the data basis for this analysis in the form of the distribution of task contents across the course of the scenarios.

Table 8.36

Distribution of Items Across All Scenarios According to the Contents and the Positions Within the Scenarios

	Position within scenarios										
Material	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Written language	5	0	5	41	5	8	0	29	15	0	7
Spoken language	32	32	11	0	0	6	58	9	4	0	0
Pictures	11	9	27	13	19	6	11	17	0	0	0
Videos	17	17	14	4	20	41	0	10	0	1	0

Table 8.36 shows the absolute position within each scenario (i.e., numbered from the beginning scene through to the last). However, the number of the scenes varied between scenarios (between eight scenes for CK, KL, FB, and MM; nine scenes for BS, CP, and R; and eleven scenes for RF). Prior to any correlative analysis, the position within each scenario was recoded according to the maximum possible number of scenes. Thus, the positions in the scenarios with eight scenes were weighted by 1.375 (positions based on nine scenes were weighted by 1.222) in order to standardize the maximum possible range of scenes to 11. If the positions were not weighted, the correlations at the high end of the item position would have relied on a reduced number of targets. Thus, weighting the positions according to the maximum number of scenes seemed to be the best approach.

As it can be seen in Table 8.36, the task contents were distributed unequally across the scenes so that any analysis based on these raw data would be biased by the influence of different task contents. The contents showed different levels of accuracy as did the different modality domains. The modality domains were unequally represented within one scene (i.e., the number of items per modality domain varied between one and six). Therefore, the means of the previously applied cells (see Table 8.32) were subtracted from the score on each item in order to account for the mean accuracy level.

The resulting overall correlation of the mean accuracy per item with the standardized position of the scene within the scenarios was around zero with *r* = .047 (N = 182; n.s.). Looking at the single scenarios, the pattern of correlations was equivocal. Table 8.37 presents the correlations for each scenario, showing mostly positive correlations. The scenarios Christop (CP) and Hannah (HR) showed a small negative correlation (*r* = -.116 / -.099, respectively). Mean Performance in the course of scenario Friedrich (FB), however, correlated substantially with the item position (*r* = .465, N = 182, p < .01). Since the level of accuracy was controlled for, this result could not be an effect of the order of scenes within the scenario.

Table 8.37

Correlations of Item Position and Performance for the Single Scenarios

	Scenarios / Target persons								
	Overall	**1 (RF)**	**2 (BS)**	**3 (CK)**	**4 (CP)**	**5 (KL)**	**6 (FB)**	**7 (HR)**	**8 (MM)**
Pearson *r* (performance with item position)	.047 (.020)	.026 (.018)	-.002 (.007)	.049 (-.034)	-.116 (-.158)	.035 (-.042)	.465** (.458)	-.099 (-.121)	.044 (.030)

Note. ** p < .01
correlations not corrected for the mean accuracy level in parentheses

These results generally suggested no substantial influence of the course of testing on the performance within the scenarios. Thus, there was no evidence for an influence of accumulated knowledge on performance. Only performance in the scenario Friedrich was significantly and positively correlated to the item position. The reason for this finding could not be clarified at the moment. However, to assess the influence of the order of scenes (i.e., information from different task material and background information) within one scenario on performance more profoundly, experiment variations would be necessary that vary the order of the scenes between groups.

Relationship to Long Term Memory

The long term memory task was based on the information contained in the scenarios. The long term memory task differentiated between memory for the information included in the task material (LTM_mat) and memory for the background information provided in the introductions to the scenarios and the scenes (LTM_back). A higher performance could point towards a more profound information processing of the respective information in the scenarios. Thus, if higher long term memory of only the task material was associated with better performance in the scenarios, this would carefully point towards perception based bottom-up processing as a determinant of social understanding abilities (Bless et al., 2004; Buck, 1983). If performance was related to the long term memory of the background information, this would suggest knowledge based top-down processing (Bless et al., 2004; Buck, 1983).

To examine this research question, two long term memory scales were constructed. One represented a general scale related to the information of task materials (i.e., LTM_mat). For example, one item required the identification of the correct wording within a written language scene. Another scale included items related to the background information contained in the scenarios (LTM_back). For example, the recall of the number of children a target person has was requested. Table 8.38 presents the reliabilities of the long term memory scales, the correlations within the long term memory scales and with the content related social understanding scales.

Table 8.38

Reliabilities of the Long Term Memory Task and their Correlations with the Social Understanding Scales

	Cronbach's alpha (Item number)	***r* within LTM $r_{mat / back}$**	**SUv**	**SUa**	**SUp**	**SUf**	**SUps**
LTM_mat	.718 (35)	.510** (.344**)	.155* (.120)	.156* (.063)	.169* (.063)	.145 (.097)	.107 (.050)
LTM_back	.527 (13)		.098 (.089)	.094 (.026)	.082 (.008)	.088 (.050)	.011 (-.025)

Note. N = 181; *p < .05, correlations with residuals of LTM (social memory tasks partialled out) in parentheses
LTM_mat = material-related long term memory; LTM_back = background information-related long term memory, su = social understanding, v = written language contents, a = spoken language contents, p = pictorial contents, f = video-based contents, ps = personality ratings

The scale reliabilities were reasonable for the material related long term memory scale (35 items). However, the reliability of the background related long term memory scale was not sufficient, but the scale was only based on 13 items. Both scales were highly intercorrelated with $r = .510$. Focusing on the correlations between the social understanding scales and the original long term memory scales, there was a small effect for the long term memory scale based on the material related information (*r* between .107 and .169). The correlations with the background information related scale were consistently smaller and not significant. However, the long term memory scales were related to the social memory tasks ($r = .148$ - .657); this correlation could be attributed to the common variance of social memory and understanding. As expected, controlling for the variance of social memory resulted in a substantial decrease in correlations between long term memory and the social understanding scales (see parentheses in Table 8.38). Thus, no evidence was provided whether performance in the social understanding tasks was determined by a more profound information processing of either the material related scale or the background information. In any event, it is certain that performance was determined to a large extent by interpretative demands as was intended by the task constructions and supported in the present analysis.

8.5 Summary and Discussion Study 2

The leading aims of Study 1 and 2 were the construction and validation of a test battery of social intelligence, the SIM. A first version of the SIM was applied in Study 1 based on a sample of 126 university students. After Study 1, item formats, the number of items, and presentation and answering times were adjusted according to the empirical results and subjects' experiences during testing. Study 2 was particularly directed at the verification of the results obtained from Study 1. However, limitations exist when comparing the two studies due to the fact that different sample characteristics such as age distribution and occupational range existed. Moreover, some task modifications or extensions were not based on empirical results, which further limited the comparison between the studies (e.g., the change of the rating format of the social understanding tasks from six- to seven-point scales). Aside from these limitations, Study 2 replicated, extended, and partly contradicted the results from Study 1. The following section will briefly summarize the main findings. A discussion of the necessary steps for further test development will be resumed in the final Chapters.

Test Construction and Psychometrics – Research Question 1A

Social Understanding Tasks

The aim of task modification after Study 1 included: the enhancement of heterogeneity of the tasks (i.e., more target persons), the balancing of the taxonomic elements within the tasks, a reduction and a standardization of testing time per scenario across the sample, the use of only one rating format (i.e., seven-point rating scales), and an increase of the number of items in order to improve the reliabilities. The test modifications in Study 2 realized the intended changes to a large extent. One problem, however, was the testing time which still varied between the test takers. However, the extent of variation could be reduced.

The screening procedures and the investigation of the psychometric properties of the data mainly supported the task modification steps. Problems with missing values or distributions did not occur. Furthermore, the reliabilities of the final scales were sufficiently high and the scales were highly intercorrelated. In Study 1, the correlations were rather small between the personality ratings and the content related social understanding scales (r = .078 - .127). Contrarily, all social understanding scales showed high correlations with the personality ratings in Study 2 (r = .472 - .645). In Study 1, the rating scales were based on 6-point scales for the content related scales and on 5-point scales for the personality ratings. Whether this change contributed to this finding, however, could not be clarified.

Problematically, the number of items showing a negative item-total correlation was large for most of the scales. Surprisingly, the same analysis relying on the more specific cells of the design generally showed a relatively larger number of items with sufficient item-total correlations. This finding suggested that the concept of homogeneity (and the application of Cronbach's alpha as an indicator for the reliability) might not be an appropriate concept for the social understanding tasks. The tasks were constructed to contain heterogeneous elements. Thus, an alternative method to determine the reliability of the scales should be considered in the future (e.g., retest reliability).

The standard deviation of the social understanding scales seemed to be restricted compared to the standard deviations in Study 1. The scale range was increased in Study 2 (i.e., from a 6-point to a 7-point rating scale) and the means in Study 2 in comparison to Study 1 reflected this change. However, the standard deviation of the tasks remained at about the same level as in Study 1 (.364 - .486 versus .389 - .472, respectively for Study 1 and 2). The origin of this finding was not clear since also the sample was more heterogeneous.

Social Memory Tasks

The problems encountered in Study 1 regarded too short presentation and answering times which resulted in a large number of missing values in the pictorial and the video based tasks (SMp2 and SMf). Due to the resulting small amounts of items, the reliabilities of the scales were sometimes low and one task could not at all be used in the analysis (i.e., SMp2). Additionally, the social memory tasks applied in Study 1 showed meaningful (but undesired) reasoning and speed requirements as reflected by high correlations to the BIS Reasoning and Speed domain. This was in part attributable to the short presentation and answering times and thus, to unwanted speed variance. Task modifications in Study 2 involved the extension of presentation and answering times, or, in turn, the reduction of the relative number of stimuli. Additionally, the number of items was increased in order to enhance the reliabilities.

The analysis in the present study showed that the task modifications succeeded only to a limited extent in improving the psychometric properties of the social memory tasks. It was proven that the number of missing values could be reduced substantially, however, the reliabilities of the final scales improved only slightly and were still not sufficient. Nevertheless, the tasks showed consistent within-domain correlations except for the second auditory task (SMa2) which also exhibited the lowest reliability coefficient.

Social Perception Tasks

The social perception tasks utilized in Study 1 did not show consistent within-domain correlations. No particular psychometric weaknesses occurred. In Study 2, one task was added for each content domain. The aim for the task development was to assimilate the underlying task requirements in terms of complexity. The analysis of the psychometric properties in Study 2 showed sufficient reliability coefficients and no severe distribution problems were encountered. The newly developed tasks based on pictures and videos (i.e., SPp2 and SPf2) showed increased difficulty which pointed towards the desired result of complexity. The resulting within-domain correlations were larger than those in Study 1, which provided some evidence that the task modifications seemed to have succeeded. However, controlling for the speed baseline variance resulted in a reduction of most of the correlations so that no coherent ability domain was supported by data.

Validity Evidence

Structure of Social Intelligence – Research Question 2A

The internal structure of social intelligence as assessed by the SIM was examined by correlational and confirmatory factor analysis. Similarly as in Study 1, the analysis supported two correlated ability factors of social understanding and memory. However, social perception could not be established as a coherent ability domain and was subsequently omitted from analysis. The factor intercorrelation between social understanding and memory was small but meaningful (r = .20) and fit statistics from confirmatory factor analysis favored the model with two correlated over two uncorrelated factors. A χ^2-differences test also revealed preference of the structural model over a hierarchical model with a general social intelligence factor. This finding contradicted the results from Study 1 and raised doubts about a higher order social intelligence construct. In Study 1, the correlation between social understanding and memory was substantially larger with r = .35 and a hierarchical model was supported by data.

Data fit statistics also supported a model establishing a faceted structure of social intelligence with two facets cross-classifying two operations (i.e., social understanding and memory) and two contents (i.e., language-based and language-free contents). However, the model showed some weaknesses with regard to the loading pattern on the content related factors which was rather heterogeneous. Thus, evidence for the validity of a content related ability facet was weak if at all existent.

Convergent and Divergent Construct Validity – Research Question 2B and 2C

The present study could not prove the convergent construct validity of the social intelligence tasks with the PONS as a measure of nonverbal sensitivity. However, correlational and confirmatory factor analysis supported the divergent construct validity from the construct of academic intelligence. Confirmatory factor analysis supported separable ability factors with. The large correlations between social memory and BIS-Reasoning and the BIS-Speed domain were reduced which pointed towards being on the right way for task modifications after Study 1. By replicating the results from Study 1, social intelligence proved the structural independency from academic intelligence. In this respect, the structural model of social intelligence with two correlated ability factors was supported when only the residuals of the social intelligence indicators with BIS variance partialled out were applied.

In extension to Study 1, the data fit supported a combined faceted model of social and academic intelligence. The social intelligence tasks loaded on the operative ability factors BIS Reasoning and Memory and on a separate “social contents” factor, complementing the existing differentiation of the BIS into verbal, numerical, and figural-spatial abilities. The content factors were positively intercorrelated. The operative factor intercorrelations were rather low (r = .08), which was interpreted as meaning that no common higher order general intelligence factor for social and academic abilities could be supported. To integrate the findings, a combined faceted model is presented in Figure 8.20 showing the two additional “social content” cells classified into the faceted BIS-Model.

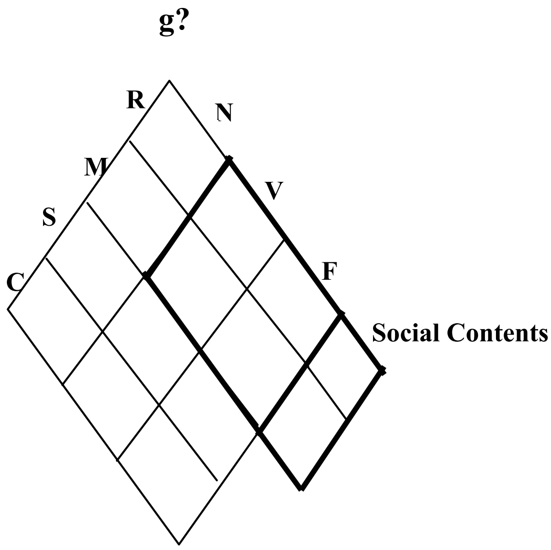


Figure 8.20

Combined Facetted Model of Social and Academic Intelligence with Social Intelligence as an Additional Content Domain

Note. R = Reasoning, M = Memory, F = Figural, C = Creativity, S = Speed, V = Verbal abilities, F = Figural abilities, N = Numerical abilities
in boldface are cells applied in the combined facetted model of Fig. 8.13

By replicating the analyses in Study 1, the present study also provides evidence for the divergent construct validity of the social intelligence tasks with personality traits. Furthermore, no substantial correlations with self-reported social and emotional skills were observed (research question 3A).

Further Findings

Gender Differences – Research Question 3B

The findings from Study 1 and 2 complemented and partly contradicted each other. Regarding the overall gender differences for the social intelligence tasks, a few effects were observed in both studies favoring female subjects. Contrary to Study 1, male subjects tended to show better performance than female subjects in the social understanding tasks, however, this effect was not significant.

Regarding the interaction between the subject and the target gender in the social understanding tasks, the findings from Study 2 clearly contradicted the results from Study 1. In the first study, an analysis of variance showed a significant main effect of the targets' gender with a higher mean performance in scenarios related to female targets. Furthermore, a significant main effect of the subjects' gender occurred with slightly better performance from female subjects.

In Study 2, an analysis of variance revealed only a main effect of the targets' gender showing a higher mean performance in scenarios related to male targets. However, no main effect of subjects' gender was observed. Moreover, by relying only on those targets that were also applied in Study 1, a different pattern of results emerged showing a comparably lower mean performance in scenarios related to female targets. This finding could be attributed to some extent to a rise in mean difficulty for the scenario Katharina (SU_KL) and to the adding of the two "female scenarios" (SU_CK and SU_HR) which were among the most difficult ones.

Scoring Methods – Research Questions 3C

One additional domain of exploratory research concerned the exploration of alternative scoring methods (i.e., group consensus scoring and correlations-based scoring). Study 2 showed that correlations between group consensus and target scoring were difficult to interpret. The correlation size depended on the item difficulty which usually represents an obscured effect not visible when aggregated scales are applied. Moreover, it was demonstrated that the bivariate distributions between target and group consensus scoring revealed a curvilinear relationship under certain conditions. This occurred when items were regarded as showing the same target answer, the same mode answer in the sample, and a minimum of a one-point deviation between the two.

Correlations-based scoring was applied as an alternative scoring procedure for the social understanding tasks. This method is originally intended to compensate for the different use of rating tendencies of both judge and target. The present study revealed one predominant weakness of this scoring method, namely, very low reliabilities. Moreover, the method as presented in Snodgrass (2001) did not consider the analysis of reliability coefficients because only one score is conventionally calculated per subject. The validity results of the social understanding tasks based on this scoring method showed no consistent inter-scoring correlations (i.e., with the target scoring scales) and low intra-scoring correlations. One other substantial change in validity results was observed in the correlational analysis of the divergent construct validity with personality traits. Correlations which were small and negative in sign (e.g., with Extraversion and Agreeableness) were still small but positive when correlations-based scoring was applied. Whether this finding could be attributed to the use of rating scales could not be clarified at this point.

Effect of Similarity on the Performance in the Social Understanding Tasks – Research Questions 3D

Cronbach (1955) established the idea that the assumed and the real similarity between judge and target contributes to the accuracy in social understanding tasks. In the present study, indicators of the assumed and the real similarity were assessed. The amount of explained variance by these indicators was generally low except for the scenario Renate (SU_RF) with $R^2_{adj} = .081$. Surprisingly, the prediction was enhanced when the perceived sympathy for the target person was added. However, for some targets, the effect of sympathy and also of the assumed similarity was negative suggesting that performance decreased when the targets were more sympathetic or more similar. To provide a careful post-hoc explanation, this finding pointed towards a reduced accuracy in information processing because the assumed similarity and / or the perceived sympathy suggested the use of heuristics to accomplish the social understanding tasks. However, this result was only valid for some of the targets and above all suggests that the number of targets should be increased in order to balance these diverse performance determinants.

Faceted Structure of Social Understanding – Research Question 3E

The social understanding tasks applied in Study 2 allowed the investigation of the structure of the social understanding tasks as postulated in the faceted design with cross-classifications of contents, modalities, and settings. No comparable analysis was possible in Study 1 since the items contained an imbalanced taxonomy.

Prior to the final analysis of the structure, the reliabilities of the cells of the faceted design were analyzed and showed comparably high reliabilities regarding the reduced number of items in the cells and more items with positive item-total correlations resulted (see above). This finding suggested that items in one cell of the design were more homogeneous than items within the content related scales as applied throughout most of the present study.

Confirmatory factor analysis of the faceted structure supported the modality facet with a differentiation into one factor related to the judgment of emotions and cognitions, and one factor related to the judgment of relationships. The loadings on the factors were stable and both factors were consistently intercorrelated across all models. Analysis furthermore supported the differentiation into a factor with loadings of all language-based cells and one with loadings of language-free task contents (i.e., the χ^2-differences test showed better fit for the two-factor instead of a one-factor solution). However, the loadings were heterogeneous on the content factors so that an interpretation of the factor was hardly possible. The faceted structure could be maintained when common target variance was controlled for. Finally, a hierarchical structure was also supported by the data suggesting the existence of a higher-order social understanding factor.

Examining the Process Underlying the Social Understanding Tasks – Research Question 3F

Two statistical approaches were applied in the present study to examine the processes contributing to the accomplishment of the social understanding tasks. One approach focused on the level of performance during the course of the scenarios. The underlying assumption suggested that the accumulation of knowledge about the target person would result in an increase in accuracy. This assumption was not supported by data; the mean performance did not rise along with the course of the scenarios except for the scenario Friedrich.

The second approach examined the correlation of performance in the social understanding tasks with the long term memory of information contained in the scenarios. In this instance, a higher long term memory of the task material would point to a more profound information processing of the material-related contents and thus suggest that perception would play a larger role (i.e., bottom-up processing; see Bless et al., 2004; Buck, 1983). A higher correlation with the long term memory of background information about the targets and the situations would point to a larger role of knowledge in the accomplishment of the tasks (i.e., top-down processing, see Bless et al., 2004; Buck, 1983).

Results showed no evidence for a relationship of long term memory of background information with the performance in the scenarios. Only small correlations were discovered for the performance with long term memory of task material. Controlling for the common social memory variance resulted in a reduction of the correlation size. Thus, neither pure bottom-up, nor top-down processing determined the performance in the social understanding tasks. Conforming to the underlying definition of social understanding, this finding, however, strengthened the role of interpretation of the social cues for task accomplishment and thus pointed to a match of the intended measurement construct and the task requirements.

9 General Discussion

The superordinate foci of the present work were threefold. One focus was aimed at establishing the theoretical foundations of social intelligence as a multidimensional cognitive performance construct. The second focus concerned the elaboration of the methodological foundations relying on already existing measurement approaches and empirical studies. Derived from these considerations were the principles of test development for a new test battery of social intelligence, the SIM. The third focus regarded the investigation of the psychometric properties of the newly developed tasks and the construct validity of social intelligence as assessed by the SIM.

The subsequent Chapters will discuss the pivotal issues with respect to the aforementioned foci and therefore, elaborate upon the differentiation into the theoretical foundations (Chapter 9.1), the methodological foundations and test construction principles (9.2), and the empirical findings with a main focus on the psychometric properties and construct validity (9.3).

9.1 Theoretical Foundations

The present work conceives social intelligence as a cognitive ability construct. It explicitly distinguishes between this cognitive abilities' approach and approaches that rely on behavioral aspects. The conceptual foundation for the present work was the performance model of social intelligence (Weis & Süß, 2005; Weis et al, 2006) and the framework of socially competent behavior (Süß et al., 2005). The performance model of social intelligence represents an integrative attempt to subsume and classify theoretical and operational definitions extracted from the literature (see Chapter 4.3.1 and 4.3.2). It originally represented a structural model distinguishing between social understanding, social memory, social perception, and social creativity as the cognitive ability domains. The model was modified in the present work by adding a hierarchical assumption in terms of a higher-order social intelligence factor. The classification of the cognitive ability domains overlaps with established ability domains in models of academic intelligence (Carroll, 1993; Guilford, 1967; Jäger, 1982). Moreover, the relevance of this differentiation also finds support in classifications from social cognitive psychology. In their attempt to disentangle the social cognitive processes, Bless et al. (2004) also distinguish between perceptive functions (i.e., social perception), interpretative processes (i.e., social understanding), and storage and retrieval functions (i.e., social memory and creativity). Both psychological disciplines, namely differential and social psychology, assign a special role to social knowledge. In definitions of social intelligence, social knowledge is still vaguely defined. It is assumed that any knowledge construct includes not only cognitive requirements, but is also strongly influenced by culture and the learning environment. Bless et al. (2004) conceive social knowledge as having an impact on all remaining cognitive processes by, for example, guiding the encoding of cues into existing categories and the top-down controlled interpretation of social cues. The subsequent test development focuses only on the domains of social understanding, memory, and perception. Some more detailed considerations will be presented in Chapter 10 about possible extensions of the SIM by the inclusion of tasks of social creativity and social knowledge.

Complementary to the aforementioned cognitive operations, further taxonomic principles visible in definitions and operationalizations of social intelligence and in related disciplines such as the psychology of emotions or social cognitive psychology are identified. These taxonomic classifications are related to *(a)* the queried modalities (i.e., the output of social cognitive operations such as judgment of emotions or personality traits), *(b)* the contents or social cues (i.e., the input that the judgment relies on such as body language or tone of voice), *(c)* the settings (i.e., the context conditions or the type of setting), and *(d)* the targets (i.e., the person or situation as the object of judgment; self vs. others; familiar targets vs. strangers). These considerations do not claim to be exhaustive, as this would not be possible. Both the number and types of taxonomic facets and the number and types of elements within one facet cannot be overlooked. For example, classifications of emotions in the literature and the associated empirical studies are manifold (Ekman, 1999; Lazarus, 1991; Roseman, 2001; Scherer et al., 2001). Comparably numerous are the classifications of discriminable personality traits. Consequently, the guiding question underlying the present and any further taxonomy must be that of the relevance and significance of the differentiations for assessing social (or emotional) intelligence. To accomplish these requirements, the taxonomic elements must be discriminable and sufficiently broad to allow adequate operationalizations and identification by the subjects, and to permit a preferably representative assessment. Another question concerns the implementation of the taxonomic elements in measurement instruments. In this respect, a differentiation of taxonomic elements is only useful when it can be realized in the intended measurement approach. For example, assessing the ability to judge emotions in familiar persons (e.g., family members) is hardly possible in group testing situations because every subject would have to bring a family member to the assessment; or the test administrator would have to accompany every subject to his or her home.

The benefits of an underlying theory-based taxonomy are self-evident (see also Cattell, 1987). It allows a more systematic and broad comprehension of the respective construct and consequently a more methodical assessment. It permits the classification of existing measurement approaches and thus a more profound interpretation of empirical findings. Taxonomic considerations can be applied as a methodological tool to balance method related variance within psychological tests. At the same time, it provides the foundation for a potential faceted model of the respective construct which allows the disentangling of different variance sources and the identification of meaningful ability domains besides operative ability factors.

9.2 Methodological Foundations and Test Construction Principles

To provide the methodological foundation for the subsequent test development, existing measurement approaches and the surrounding methodological problems and challenges were addressed with a focus on the class of cognitive ability tests. Existing approaches were discussed in the light of the methodological shortcomings and the resulting validity evidence. The resulting empirical problems, above all, concern low reliabilities and the lack of convergent and divergent validity evidence. In light of these considerations, the most striking problems are identified as the use of artificial and decontextualized item material, the use of only written language contents, a lack of theoretical a-priori considerations and, consecutively, a mismatch between the purported measurement construct and the actual task requirements.

During the course of test development, several decisions were necessary which partly relied on the aforementioned criticisms of existing approaches and which are more or less debatable. These concern the test principles of all social intelligence tasks (i.e., the application of genuine task material, the differentiation into the different content domains, and the selection of response formats) (Chapter 9.2.1) and specific questions surrounding the social understanding tasks (i.e., the application of a scenario approach and the application of the target scoring procedure) (Chapter 9.2.2).

9.2.1 Test Principles of all Tasks

Genuineness of Task Material

One relevant claim for all newly developed tasks was the application of genuine task material. The most striking advantages of genuine compared to posed or artificially produced material are the ecological validity and the availability of scoring standards for the social understanding tasks. With respect to the ecological validity, the task material in the present test development represents a variety of persons and situations. Any test battery based on posed material does not seem to be capable to comprise a comparable variety of contents. Any fictitious story or script cannot reflect the diversity of real life that often provides unexpected events; it is the unexpected contents that occur in real life that could hardly be included in scripts. Although professional script writers have certainly learned to account for this problem, it can be assumed that posed or artificial task material tends to favor "typical" situations that occur in real life. With respect to the honesty in providing truthful answers to questionnaires and to the recorded scenes, this implies several benefits and problems that are addressed in a later section when the target scoring procedure is discussed.

Other disputable issues refer to the efforts of material sampling for the target persons and the consequences for the test constructors. The target persons, particularly those included in the social understanding tasks, were accompanied for a certain period of time (i.e., between a few hours and two days) in both private and public situations. They were recorded by digital camcorder, voice recorder, and photocamera. Additionally, some were required to fill out questionnaires, provide answers to the recorded scenes, and hand out text material. The resultant questions concern the authenticity of the exhibited behavior and the honesty of the information provided. With respect to authenticity, the experience during the recordings clearly suggested that targets get used to the fact that they are being observed (e.g., video-taped) after a certain time after the beginning of the recordings. This was confirmed by the colleagues who were responsible for the recordings and of whom one was always acquainted to the respective target. The consequences of relying on genuine compared to posed or artificial task material for the test constructor are distributed on different points throughout the test construction process. Prior to the recordings, the exertion of writing scripts represents a greater effort compared to "simply" selecting target persons and accompanying them over two days. After the recordings, the work on the sampled material represented the greater effort including the assessment of the target information (only in social understanding tasks), the selection of the most appropriate scenes out of all of the recordings, the matching of the material contents with the taxonomic demands, and the editing of the scenes in order to achieve adequate item formats.

One disadvantage of genuine task material lies in the quality of the task material. The quality was sometimes limited in the present test construction because no artificial means were applied to modify the situation (e.g., the setting up of floodlights). However, the quality of most of the material was good and subjects were instructed to ignore any restrictions in quality.

Differentiation of Task Contents

The taxonomy underlying all tasks was intended to help balance the situative input and the contents of the task material (see Chapter 6.2.1). The latter varied in terms of the applied task material (i.e., written and spoken language, pictures, and videos). Spoken language contents did not include pictorial information, and videos did not contain an audio stream. This classification originally served as a methodological tool to balance method related variance. The systematic variation of contents was also conceived as a variation of the social cues displayed in the task material. However, it must be questioned whether this classification represents an appropriate differentiation of social cues (occurring in real life) or whether it is rather artificial. Several classifications seem sensible in real life, however, they overlap to a certain extent and do not allow unequivocal assignment of the type of cue to the type of content. For example, social cues can diverge into language-based and language-free (nonverbal) cues while language-based cues can rely on written or spoken language. At the same time, spoken language also contains cues that are not intrinsically related to the language but to the way of using the language itself (e.g., the intonation). This fact was used in the auditory social perception task (SPa2) where artificial (senseless) spoken sentences were applied and subjects had to identify the correct emotion relying only on the “language-free” cues such as intonation. In turn, originally language-free cues contained in pictures or videos without including an audio stream contain totally different cues such as postures, gestures, mimic, etc. which differ from the language-free cues in spoken language. Furthermore, the different cues related to the task material also differ in terms of the approximation to cues occurring in real life. Written and spoken language cues occur in letters, emails, and telephone conversations as typical ways of communication. Photographs are probably the most unusual social cues compared to the remaining task contents. Moving pictures (i.e., videos) without auditory information are probably seldom relevant because conventionally, people can listen to communication while viewing the person. However, situations are possible to think of where people watch others from a certain distance so that they cannot listen to the conversation but still interpret the nonverbal cues they see from a distance.

This fact notwithstanding, it seems that it is the combination of nonverbal cues (i.e., from the tone of voice to postures and gestures) and language-related information which is one meaningful aspect in human interactions. This combination, however, was not included in the design of the SIM and the consequences cannot be determined. It is certainly true that, in this respect, the sum of the parts does not equal the whole. It cannot be clarified whether performance particularly in the social understanding tasks is enhanced when combined task contents are applied. Contrarily, Archer and Akert (1980) claim that providing all possible cues may be confusing for the subjects because cues sometimes stand in contradiction to each other. Their results, however, showed that performance was highest when the full material was provided. At any rate, the results in Study 2 did not reveal a particular bias except for the fact that written language material was the hardest to interpret in the social understanding tasks. In the end, an answer to this question requires an empirical investigation.

Response Format and Item Selection

Some conceptual weaknesses of the present work concern the procedure of the selection of adequate response formats for the social understanding tasks, the applied response formats of all of the tasks, and the process of item selection.

Regarding the selection of response formats for the social understanding tasks, Study 1 was directed at empirically identifying scale properties depending on the response formats (i.e., ratings-based, multiple-choice, and open-ended formats) used. However, the first task version did not systematically vary the response formats across different contents and modalities. Moreover, the number of items per format was not balanced. Thus, separate scales related to the different formats could not be built so that the question of format selection could not be answered empirically only. The actual procedure compared scales based on all item formats to scales based on rating format only and found comparable reliability coefficients and larger within-domain correlations for the ratings-based scales. Additionally, the decision to apply only the rating format in the subsequent task modifications was also based on the question of economy of testing and test construction. However, together with the task modifications of the social memory tasks, this decision resulted in different response formats in all three operative ability domains. In Study 2, rating formats (social understanding), multiple-choice and open-ended formats (social memory), and response latency formats (social perception) were applied. As MacCann (2006) demonstrated, the use of different response formats resulted in a drop in correlation size. It can be speculated that the lower correlation between social understanding and memory (on a latent level) in Study 2 was, in part, due to different response formats.

With respect to the last issue, the selection of adequate items for the final scales in Study 1 was based on the item-total correlations (besides conceptual considerations related to the taxonomy). The resultant Cronbach's alpha coefficient consequently represented an overestimation because the scale was optimized for the respective sample. Problematically, the modifications of nearly all tasks in Study 2 did not only refer to the application of the selected items, but frequently included further changes such as the response format, different distracters, and different underlying item material. Therefore, the effect of only item selection could not be examined in Study 2 and the item-total correlations could not be compared to those in Study 1. This could explain the partly large number of negative item-total correlations and would require one more step in test development, namely the confirmation of the present item selection in another sample.

9.2.2 Test Approach of the Social Understanding Tasks

Selection of the Test Paradigm

Different from the test approaches for the social memory and perception tasks, the social understanding tasks represented an unparalleled attempt, without a specific predecessor in the literature, which combined a scenario approach with the so-called postdiction paradigm. A scenario presented a target person in his or her natural surroundings based on scenes of all four types of task material. The advantages of this approach lies in the opportunity to introduce context information in an efficient way. The introduction of context information would not be accomplished for a comparable number of items if each item referred to a different background. In the present approach, one item refers to various context information. These can be differentiated into the specific item background related to the respective scene, the overall background related to the target person, and the general social background (e.g., the biographical information about the targets or all other information about the life of the targets and the situations).

The effect of the use of such rich context information on the performance beyond the influence of the information provided by the task material could not be clarified in the present work. An open question remains that addresses a possible anchoring effect of specific information and the activation of knowledge structures about the target persons or the situations. Anchoring effects concern, for example, the elicitation of expectations by the instructions to specific scenes. It is important that these instructions are formulated as neutral as possible without suggesting the target responses to subsequent items. At the same time, the instructions are supposed to direct the subject's attention to the relevant cues in upcoming scenes. It cannot be clarified whether the balance between neutrality and attention control was accomplished in the present work and this question must be subject to further investigation.

It is unclear whether and how much subjects made use of existing knowledge structures in judging the mental states of the targets. The task definition identifies that only the interpretation of the task material determines performance. However, the situations in the scenarios or the target persons possibly elicit the memory for one's own experiences and associated knowledge structures. If this happens and subjects do not explicitly exclude the influence of such knowledge, potential top-down processes may influence information processing. In this respect, stereotypes represent an extended type of top-down processing relying on heuristic information processing. Stereotypes may also be elicited by certain general person characteristics of the targets (e.g., gender or religion stereotypes, stereotypes of a specific profession such as a bartender, etc.). The designs of the studies in the present work do not allow the investigation of these effects. Some ideas about future research questions and designs will be provided in Chapter 10.2.

Target Scoring Procedure

The scoring of the tasks of social understanding or related social or emotional abilities is subject to controversial discussions (Cronbach, 1955; Legree et al., 1995; Mayer & Geher, 1996; Schulze et al., 2007; Tagiuri, 1969; Wilhelm, 2005). Several scoring procedures are suggested (i.e., target scoring, group and expert consensus scoring, standards-based scoring). The problems of group consensus scoring were already discussed in detail (see Chapters 5.2.4.2 and 8.4.4.3) and shall not be mentioned again. Standards based scores could not be applied in the present test approach. The chosen scoring in the present work was target scoring, however, this procedure is also debatable. The most crucial concerns in the literature and from the present work are the validity of the target information about their mental states, the accounting of Cronbach's (1955) original critique of the influence of stereotype accuracy on the target score, and the specific algorithms to calculate the target score.

Regarding the validity of the target information, this represents a serious problem encountered in the discussion about target scoring. It is frequently questioned whether target persons are capable of reporting their true mental states because these may be too complex or too difficult to communicate or perhaps may be deemed socially undesirable (MacCann et al., 2004; Mayer & Geher, 1996). Some researchers, however, also report that targets are more precise in judging their mental states compared to external observers and, at the same time, not as precise in judging their own behavior (Spain, Eaton, & Funder, 2000 as cited in O'Sullivan, 2007). In the present work, it was attempted to account for single imprecise or intentionally dishonest answers throughout the process of item construction and selection by sorting out items where answers contradicted with what was obvious in the recordings of the respective situation or with any other information available. However, it could not be clarified entirely whether, at some point, the targets were not providing the "right" answer about their mental states. Therefore, necessary future directions concern the application of peer and expert ratings to validate the target answers. Moreover, necessary test documentation needs to archive which are the relevant cues within the task material that support the target answer (see Chapter 10).

Cronbach's (1955) critique of the target scoring procedure referred to the influence of separate accuracy components that constitute the target score and that bias a simply calculated difference score. The most important component represents the stereotype accuracy (i.e., the degree to which a judge predicts the mean answer of a group of targets). According to Cronbach, the single components are influenced by the interaction of the assumed and the real similarity between judge and target. The present scoring procedure did not account for this alleged problem. Since the targets were judged on different items, it was not possible to conduct the suggested componential analysis (Cronbach, 1955) that decomposes the variance of the target score into the different types of accuracy (i.e., stereotype and differential accuracy; see Chapter 5.2.4.1 for a detailed description). Thus, it could not be clarified whether or to what extent the present target score was determined by the Cronbach components. Analysis showed that the real and assumed similarity or the perceived sympathy was not meaningfully correlated with performance in the scenarios. Only some specific effects related to specific target persons could be identified. This analysis certainly does not imply the focus of Cronbach who was only regarding the use of rating format and not the general idea of assumed or real similarity. However, without relying on empirical data, some arguments can be considered that seem to put Cronbach's critique into another perspective: *(a)* subjects were instructed to show maximum performance so that they should show a maximum strive for accuracy, *(b)* subjects have maximum resources and unlimited time available. If they had not, this would usually limit profound information processing and enhance the probability to apply mental shortcuts such as stereotypes (Bless et al., 2004; Tagiuri, 1969), *(c)* the scenarios contain unfamiliar situations and persons which should minimize a bias due to prior knowledge, and *(d)* eight targets were applied so that a potential influence of stereotype accuracy or any other heuristic information processing due to similarity, prior experience, or sympathy should be balanced in the content related scales across the targets.

A final problem that must not be ignored concerns the specific mathematical procedure applied in the present work. The target scores were calculated by the weighted differences between the subject and the target answer. The weights accounted for the position of the target answer and the resulting maximum possible deviation (see Chapter 6.2.2.1). Thus, the basis for the scoring was the target answer on the 7-point rating scale (6-point in Study 1). This answer, however, was, in turn, calculated from the original target answer which took place on the 10 cm analogous scale (i.e., divided into 100 mm for an original answer between 0 to 100; see Chapter 6.2.2.1). Consequently, for the 7-point rating scale, an original target answer of 0 to 14 received a score of 1, from 15 to 29 a score of 2, etc. This procedure implied two major problems: *(a)* some original target answers lay at the border between one rating category and the next (e.g., an original answer of 14 received a score of "1"; an original answer of 15 received a score of "2") and *(b)* the extreme original answers (0 and 100) were put into the same extreme rating categories ("1" and "7") as less extreme original answers (e.g., 10 or 90). With respect to the latter point, the endpoints of the final rating scales were formulated as to refer to the extreme cases "e.g., *not at all present* or *extremely present*). To provide insight into the relevance of this problem, the original answers of four targets (CK, CP, FB, and MM) were inspected and showed that a meaningful number of original target answers are affected by the assignment of original answers to rating categories. Out of a total of 236 original target answers, 18 lay exactly on the category borders. For the example from above, this would reflect an original answer of 14 or 15. Additionally, 20 answers were one point beyond the borders (e.g., 13 and 16). However, for these four targets, only three original target answers belonged to the extreme categories without representing the true extreme answers of 0 or 100 so that this problem seems negligible.

However, it seems interesting to recalculate the scores in order to explore the impact of the present procedure. Particularly, it seems interesting to reassign the extreme categories to only the extreme original answers and to create five categories in between by dividing the original answers between 2 and 99 by 5. This would also result in a smaller number of original answers that lie exactly on or close to the category borders.

9.3 Studies and Empirical Findings

Sample and Procedures

The empirical findings in the present work are based on two main studies. Study 1 included a sample of 126 German university students (m_{age} = 21.35; *sd* = 3.06; 53.5 % females). In Study 2, an unselected sample of adults participated (m_{age} = 28.69; *sd* = 5.57; 58.8 % females). The subjects in Study 2 were heterogeneous in terms of age, education, and occupation. The total testing time, including breaks, was 10 hours for Study 1 and 12 hours for Study 2. Testing was conducted on two separate days, with half of the testing time occurring on each day. On every testing day in Study 2, the testing was organized in four blocks with a break in between each block.

One concern refers to the lack of comparability of the two samples. The samples not only differ in terms of biographical features, but also in terms of the mean performance level in the BIS scales (see Appendix D). This complicates the interpretation of the mean difficulty levels of the newly developed tasks. At the same time, it is a notable finding that the models of social intelligence structure could be replicated in Study 2.

Another concern addresses the rather long testing time per day (i.e., 5 and 6 hours, respectively in Study 1 and 2) and possibly related problems of compliance or exhaustion. In Study 2, the perceived exhaustion was assessed at four different time points during the testing (i.e., Day 1 Block 1, Day 1 Block 4; Day 2 Block 1, and Day 2 Block 4). The mean level of exhaustion rose from the first to the last block, however, it did not reach an extreme level (see Table 8.8 in Chapter 8.3). Moreover, the perceived exhaustion in one block did not correlate with the performance in the tasks of the respective block (higher exhaustion expressed by lower scores). In Block 1 on Day 1, the correlations ranged between -.140 and .008; in Block 4 on Day 1, from -.012 and .056. In Block 1 on Day 2, the correlations ranged from -.129 to .061 and for Block 4 on Day 2 from -.102 to.061.

Thus, it is apparent that perceived exhaustion did not play a role for performance which corresponded to the experiences during the testing. Only very few subjects complained about the long testing time or asked for an earlier break.

Empirical Findings

a) Scale reliabilities

One big concern represents the reliability analysis of the tasks. Cronbach's alpha was applied as an indicator of the reliability. This coefficient requires the homogeneity of the item true scores of one scale. However, the task conception is meant to be heterogeneous so that it is questionable whether really high reliability coefficients can be achieved at all. The analysis showed mostly sufficient reliability coefficients, however, all scales were optimized by excluding items with a negative item-total correlation. Interestingly, the number of items with a negative item-total correlations in the social understanding tasks was smaller when the entity of analysis was smaller (i.e., when only the cells of the social understanding tasks were analyzed). This supported the assumption that the items were rather heterogeneous when the overall scale level was regarded. In any event, future research should investigate the retest-reliability in order to provide further insight into the reliability of the newly developed tasks. Any further issues of future task modifications based on the psychometric properties of the present tasks (e.g., the lower reliabilities of some social memory tasks) are described in the upcoming Chapter 10.

b) Item difficulty of social understanding tasks

Another issue requiring some sophistication represents the problem of difficulty of the social understanding tasks. Several analyses throughout both studies yielded varying degrees of difficulties, depending on the type of scales applied, furthermore partly inconsistent findings. The most debatable empirical findings are to be repeated hereafter: *(a)* the content-related social understanding scales showed varying difficulty patterns (i.e., in Study 1, the pictorial scale; in Study 2, the video-based scale showed the lowest difficulty level), *(b)* the difficulty level related to the different targets showed inconsistent findings (i.e., scenario Katharina was among those with the lowest difficulty in Study 1 and among the most difficult in Study 2), and (c) the cell *judgment of cognitions based on pictures* showed a rather low difficulty although the remaining cells based on pictures were among those with the highest difficulty. Every single one of those findings could have been underpinned by some post-hoc explanation. However, after all, it appeared inevitable to develop a schema to deliberately estimate and adjust the item and scale difficulties. Such a schema should account for the determinants of difficulty regarding the different taxonomic elements of the social understanding tasks. First steps in this undertaking could be *(a)* the obtaining of expert ratings and *(b)* the documentation of the relevant cues within task material. This could be compared to the item difficulties in the present sample. Both the assembling of expert ratings and further design-related issues to deal with this problem will be addressed in Chapter 10.

c) Social perception tasks

Another concern that emerged during the analysis refers to the lack of convergence of the social perception tasks. The intercorrelations seemed to support a coherent construct however, controlling for the speed baseline measures resulted in a substantial reduction of correlation size for most of the tasks. Confirmatory factor analysis could also not identify a common social understanding factor. Seidel (2007) elaborated her findings concerning the social perception tasks in detail and concluded that different underlying processes contributed to the apparent lack of convergence between the social perception tasks. She differentiated between processes that only require sensory perception based on physiological channels (i.e., auditory cues such as laughter) and those that require a larger amount of processing (e.g., perception of written language cues). This idea is similar to the differentiation of Bless et al. (2004; see also Chapter 4.3.3.1) between perception and encoding as processes that address any incoming information. The authors refer to perception as the pure physiological perception process of stimuli entrance. Contrarily, encoding functions reflect the process of assigning a mental representation to perceived stimuli by the use of available social categories (e.g., assign “smile” to a perceived movement of a person’s mouth, or “man” to a perceived person). Regarding this classification, the elicited perception process depends on the applied stimulus. It could be assumed that the concreteness of the stimulus and the sensory channel determine the subsequent process. However, research is needed that addresses this question empirically.

d) Construct validity of social intelligence

Regarding the internal structure of social intelligence, the analysis in Study 2 did not convincingly support a general social intelligence factor which contradicted the findings from Study 1. A hierarchical model showed good data fit, however, a comparison with the structural model showed better fit for the structural model. This finding clearly needs replication. First and foremost, it should be accounted for the possibility that this finding was due to the different formal characteristics of the items underlying the social understanding and memory domain (e.g., ratings-based scales vs. multiple-choice and open-ended response format, untimed vs. timed administration; respectively). However, if this finding was confirmed in future research, the performance model of Weis & Süß (2005) needed to account for tthis lack of coherence.

With respect to the divergent construct validity, the present analysis clearly demonstrated the benefits of applying a sophisticated measure of academic intelligence such as the BIS-Test (Jäger et al., 1997). Eventually, the divergent construct validity of social intelligence as assessed by the SIM was demonstrated although analysis relying on specific levels of the hierarchical BIS-Model first raised doubts about the construct independence (i.e., large latent factor intercorrelations between social memory and BIS-Memory with $r = .67$).

e) Content-related ability factors in faceted designs

One pivotal question emerging from the investigation of faceted models throughout Study 2, of both the SIM and the BIS and the SIM combined, concerns the content-related ability factors. Although the faceted structure of social intelligence, and of social understanding was supported by confirmatory factor analysis, results showed heterogeneous and partly uninterpretable loadings on most of the postulated content factors. They, in times, seem to allocate "residual" variance so that a meaningful differentiation within the content facet was hardly possible. Only one language-free factor in the faceted model of social understanding showed a rather coherent loading structure. At the same time, a combined faceted model of social and academic intelligence supported a "social content" factor with heterogeneous but consistently directed loadings of the social intelligence indicators. To attempt putting these findings in context, some more issues need attention.

The social understanding tasks did not show the typical bias related to written language task material as frequently found in past empirical studies. The tasks were only marginally related to the verbal BIS content domain. This fact may be responsible for the lack of common "language" variance with social understanding tasks based on spoken language and thus, for the diverse loadings on the language-based factor of social understanding. Moreover, the spoken language contents do not only rely on language-related cues what may additionally imply a different impact. In contrast to the language-related contents, the pictorial and video-based tasks of social understanding show at least some meaningful common content variance allocated on a language-free content factor.

Regarding the findings from the faceted model of social intelligence, the social understanding and memory tasks differ in one important aspect that may be responsible for the lack of common content-related variance. In the social understanding tasks, the content-related differentiation only has an impact on the task material and thus on the provided cues. Since the task conceptualization required the interpretation of the task material, the task product is only indirectly related to the cues in the task material. In contrast, the content-related differentiation in the social memory tasks also influences the task product because the task conceptualizations required that only objectively provided information must be queried. Thus, tasks based on written and spoken language pose items only referring to the language contents (as the only objective cues).

To summarize, while the social understanding tasks apply to a distinction of different queried modalities related to a basically independent differentiation of task material (i.e., a full cross-classification is possible of modalities and material-related contents), the social memory tasks only query objective cues that directly depend on the task material (what information is queried directly depends on the material-related contents). This could explain why a differentiation into task contents was not truly relevant for the social understanding tasks so that no meaningful common content variance could be discovered with the content-depending social memory tasks.

Conclusively, the question remains which social contents were allocated in the social content factor in the combined faceted model of social and academic intelligence. This model seems to support Guilford's (1967) conceptualization of social intelligence as an autonomous content domain (i.e., behavioral contents) besides, for example, symbolic and semantic contents. However, the loadings on this factor and the factor intercorrelations with the BIS suggest that it is rather language variance that determines the social content factor. The language-based cell of social memory showed the highest loading (.78), and the factor was correlated with $r = .52$ with BIS-Verbal abilities. The remaining loadings of the social intelligence indicators were substantially smaller but all positive so that at least, weak evidence was provided that the factor was also determined by contents beyond those related to language.

f) Analyses relying on the target-related social understanding scales

The social understanding tasks were based on eight scenario, each related to one target person. The actual focus of the present work lay on the construction and investigation of social understanding scales related to the task contents to match the MTMM design of the SIM. In this original conception, idiosyncratic effects of different targets (e.g., similarity and sympathy between judge and target or different sending accuracy of targets) were all subsumed under the idea of item difficulty. Thus, the aforementioned replies to Cronbach's (1955) critique also applies at the present point in terms of the maximum performance situation induced by the tasks with unlimited resources provided.

Some analyses in Study 2 switched the focus from the content-related scales to scales that aggregated items related to the target persons. Results showed no systematic influence of the possible biases as far as they could be accounted for. However, some effects were discovered related to specific scenarios. For example, performance in the scenario Renate was predicted to no meaningless extent by indicators of assumed and real similarity. The perceived sympathy for the target person could, moreover, add significantly to the prediction in three other scenarios (i.e., Bringfried, Christoph, and Friedrich). Performance in the scenario Friedrich increased substantially with the course of testing although different accuracy levels of contents and modalities were accounted for. It is certainly possible that these were rather random effects because there are no obvious reasons for these findings.

Anyway, the most prevalent conclusions suggested that the application of as diverse and as many targets as possible seems important. The upcoming Chapter will again recur to this problem by providing design suggestions.

10 Future Perspectives

This last Chapter is intended to provide suggestions about future steps in test development and research designs capable to complement and extend to the findings of the present work.

10.1 Test Modifications

The test version of the SIM applied in Study 2 requires concrete steps of modification relying on the identified psychometric weaknesses and some debatable results concerning the internal structure and the construct validity. The upcoming Chapter will provide suggestions for test modifications based on the aforementioned considerations.

Technical Modifications

The advantages of implementing the social intelligence tasks in the experimental software WMC 0.18 are self-evident and refer to the standardization of task administration, the sampling of data directly on the computer, and the exact assessment of reaction time scores. However, one concern about the technical restrictions related to the experimental software represents *(a)* the lack of possibility to present video and audio information at the same time and *(b)* the comparably low quality of the pictures due to the constraints on data format. Therefore, it should be attempted to find means to deal with these restrictions.

Reliability Analysis of all Social Intelligence Tasks

The test approach explicitly introduces heterogeneous task material for all tasks so that the problems related to a large number of items with negative item-total correlations and partly low reliability coefficients of internal consistency do not seem surprising. It is by no means an alternative to reduce heterogeneity in order to enhance the homogeneity of the items. Apart from some specific modifications of single tasks that are also necessary because of taxonomic considerations, the breadth and representativeness of the tasks is to be obtained. Therefore, it appears outmost important to investigate the stability of the test scores in a retest investigation and thus, conform with one of the requirements on a new intelligence construct of proving stability (Süß, 2001).

Task Modifications of Social Understanding Tasks

The objectives for future task modifications of the social understanding tasks are listed as follows: *(a)* the reduction of the total test length in order to allow more economic testing, *(b)* the development of a schema to estimate and adjust item difficulties, *(c)* examining the validity of the target answers, and *(d)* considering the use of combined audio and video information to enhance real-life fidelity.

a) Reduction of test length

At present, the scenario tasks take about three hours of testing time. It is totally unrealistic to maintain this test length, both for research and applied settings. Several starting points for the reduction of test length are conceivable, for example, the exclusion of entire scenarios, the exclusion of entire scenes within one scenario or the exclusion of single items. The latter option could rely on the results of the present study by selecting items with only positive item-total correlations. However, this may result in the least reduction of test length since still the same scenes were included in the scenarios. The other options are not to be derived as obviously from the present results since no entire scenario and hardly any entire scene showed malfunction. This dilemma can hardly be solved at present and needs further considerations based on the following concerns of test modifications. However, any actions require a thorough consideration of the consequences when entire scenes or further background information is excluded from one scenario task.

b) Learning about item difficulty and c) validity of the target answers

These two objectives are combined since they rely on comparable steps. As elaborated in the discussion, several analyses point to controversial results relating to the item and scale difficulties. It appeared valuable to develop a schema to learn about the determinants of item difficulty. Therefore, peer and expert ratings appear promising, such that, at the same time, the validity of the target answers could be examined. Expert ratings should not only assemble the expert answers on the items but also investigate the experts' opinion about the underlying processes specific to task material. Additionally, the relevant cues within the material for the accomplishment of the tasks should be documented for every single item, both based on the expert statements and on the material assembled during scenario construction. The results of this documentation can be interpreted in light of the experts' and peer ratings and the psychometric properties as found in the present work.

With respect to the validity of the target answers, an investigation of this appears as a necessary step to confirm the objectivity of scoring in light of the manifold criticisms related to the target scoring procedure.

d) Application of task material combining audio and video information

It appears interesting to develop a parallel task that relies on the combined audio and video information in order to allow a comparison of the accuracy levels. Moreover, this type of task material may as well represent the most genuine situations so that the relevance for real-life (applied) settings may be enhanced in such a task version.

It is suggested throughout the present analyses and the preceding discussion that it appears promising to apply as heterogeneous and as many targets as possible in order to avoid idiosyncratic effects (see last section of Chapter 9). This suggestion contradicts to the objective to reduce the test length. A compromise accounting for both aims is certainly not easy to find based on the present approach. The introduction of context information and the maintenance of taxonomic requirements need a certain amount of task material and introductions to the subjects. One possible way of dealing with this problem could lie in abstaining from within-scenario variation of the taxonomic elements, but rather introduce taxonomic variations across the single scenarios.

Task Modifications of Social Memory Tasks

The necessary task modifications for the domain of social memory concern more specific aspects related to presentation and answering times and the consecutive item difficulties. The pictorial and video-based tasks of social memory still showed missing values located at the end of each sub-task indicating too short answering times. The bigger problem concerns the low scale reliabilities. These may also be due to the heterogeneity of task material and requirements. Thus, apart from some concrete item modification or selection based on item-total correlations and item difficulty, the social memory tasks also require a retest investigation to examine the stability of the scores. Similar to the social understanding tasks, the social memory tasks need a reduction of the test length in order to account for the standard of economy.

Task Modifications of Social Perception Tasks

The outline for possible task modifications of the social perception tasks is not as concrete as the preceding considerations since no particular psychometric problems occurred. Chapter 9 already elaborated the possible problems underlying the lack of convergence between the social perception tasks. Therefore, task modifications require more elaborate a priori considerations that account for the effect of different cognitive functions as determinants of social perception (i.e., sensory processes related to physiological channels vs. encoding processes requiring a basic processing of the stimuli). Task modifications or new development should consider these different processes so that a coherent domain of social perception or related requirements can emerge in future research.

10.2 Future Research Questions and Designs

To conclude the present work, expanding research questions and design issues derived from the theoretical and empirical parts of the present work shall be elaborated. These concern *(a)* expanding research questions related to the social understanding ability domain and the present test approach *(b)* considerations about the performance model of social intelligence as established by Weis and Süß (2005) and modified in the present work.

Expanding Research Questions Related to Social Understanding

Expanding research questions related to the ability of social understanding as assessed by the scenario tasks all refer to the investigation of the underlying processes. It is frequently argued throughout this work that the task conceptualization of social understanding incorporates processes that cannot be fully estimated at present. The scenario task rely on rich context information so that it could be questioned whether performance relies on the actual task material or the background information. An open question remains about a possible anchoring effect of specific information within one scenario and the activation of knowledge structures by specific information (top-down processing). It is unclear whether and how much subjects make use of knowledge in judging the mental states of the targets.

The task conceptualization underlying the scenarios clearly states that only the interpretation of the task material is supposed to determine performance. Results in the present work do not suggest any diverse process to be relevant. However, the scope of this analysis is restricted. Rather, these questions demand more design-oriented approaches. Therefore, some suggestions shall be made that concern possible ways of exploring the processes of the social understanding tasks:

a) Promoting and examining bottom-up vs. top-down processing:

The use of bottom-up vs. top-down processing is conceived as depending on the processing motivation and the available resources (Bless et al., 2004; see also Probst, 1982 for the differentiation between intuitive and deliberate information processing). Several experimental conditions are thinkable that promote either top-down or bottom-up processing. *(1)* Processing motivation and thus, bottom-up processing should be promoted when subjects are instructed to justify their answers (see Bless et al., 2004; e.g., "Tell me the basis for your judgment and the inferences you have made."). Moreover, the technique of thinking-aloud should elicit the same effect. Such investigations should provide helpful information about the general process and effects related to specific scenarios or scenes. Both conditions require effortful testing, and subjects can be tested only one at a time. *(2)* In turn, processing motivation is supposed to be reduced (i.e., enhancement of top-down / knowledge-driven processing) when resources are withdrawn (e.g., in a dual task paradigms or by the introduction of time limitations). Forcing people to make quick judgments enforces automatic, knowledge-based information processing (i.e., top-down).

b) Manipulation of the cues and background information

Other experimental conditions concern the manipulation of task material and background information. For example, one condition can exclude any of the background information and only provide subjects with the actual task material. Another condition can omit the task material and provide subjects with biographical background information about the targets and a situation descriptions. The latter condition (i.e., no available cues) should be capable to enforce the use of stereotypes since subjects would not know anything about the target's behavior in the respective situations and they would not be provided with specific, idiosyncratic cues.

The aforementioned experimental manipulations can be applied in a both within- and between-subjects designs. These should also allow a comparison with the standard testing condition as applied in the present study. Comparing the accuracy levels of the experimental groups probably yields more information about the concurrent processes during task accomplishment.

Modifictaions to the Performance Model of Social Intelligence (Weis & Süß, 2005) and Extentions of the SIM

Analyses in the present work carefully points to two modifications of the performance model of Weis and Süß (2005) that need replication in future studies. On the one hand, the social perception ability domain was so far not supported by data (see also Seidel, 2007) so that, without any further investigations, it cannot be justified to keep it in the model as a meaningful ability domain. On the other hand, evidence for the hierarchical assumption of a higher-order social intelligence factor was weak. This could be due to many different facts (e.g., different response formats, different queried modalities of social understanding and memory, etc). Thus, further research is needed that incorporates the aforementioned principles of task modifications. Besides the implications of the empirical findings for the performance model, in turn, the model also implies some extensions of the SIM regarding the adding of two more ability domains, that is, social creativity and social kowledge.

Social Creativity

The current test battery focusses only on three cognitive ability domains from the performance model of Weis and Süß (2005). However, the model also includes social creativity as a forth domain, sometimes also labeled social flexibility. It was defined as the production of as many and as diverse solutions or explanations as possible for a social situation or problem (Weis & Süß, 2005; see also Lee et al., 2002). Social creativity was not realized in the present test development basically due to economic constraints. Nevertheless, this ability domain is sought to be as relevant as the remaining domains, although surely social understanding plays the most prominent role. To provide an example of the practical relevance of this ability domain, imagine an unfamiliar situation where equivocal cues are provided so that an inference from the cues is not possible. Social creativity abilities can now help finding *possible* explanations for the equivocal cue pattern. The products of socially creative cognition can then, for example, guide further information seeking in order to finally achieve an explanation for the situation. Abstractly spoken, social creativity represents imagination ability. This example illustrates why this ability domain is frequently referred to as retrieval functions (Bless et al., 2004) or the flexible application of knowledge (Lee et al., 2002). It seems that in this conceptualization, social creativity hardly, if at all, depends on the available cues but rather on the ability to combine prior knowledge or experience with the information about the poblem and not necessarily, the equivocal cue pattern. Performance criterion is not the idea of a correct or incorrect solution, but rather that of quantity and diversity.

The present work extended the performance model of social intelligence by the introduction of taxonomic classifications of contents, modalities, settings, and targets. To apply this differeniation to social creativity, just as social understanding, the outcome of socially creative cognition is not directly related to the available cues (conventionally provided by the task material). What is queried in the social creativity domain (i.e., the product), are possibly complex social facts such as solutions to problems, explanations to a situations, possible future behavior, but also a possible range of emotions someone might possibly feel about a certain event. Thus, the classification of different queried modalities also counts for the social creativity domain. The taxonomic considerations related to the settings or the targets also apply one-to-one.

Consecutively, the aforementioned considerations suggest that the ability domains purportedly related to the social intelligence construct possible need some further differentiation, so far not applied explicitly. This differentiation concerns the relationship of the task material to the outcome (product) of a task. In the social memory and perception domain, the task outcome is directly related to the task material (the input), no extrapolation from input to outcome is necessary or desired (i.e., only objectively present contents can be addressed as task outcome). Contrary, social understanding and creativity reach beyond the given information so that the queried task product does no longer directly relate to the task contents.

To finalize these considerations, a potential test approach to social creativity is illustrated. According to the domain conceptalization, a test of social creativity requires the posing of an equivocal situation to the test taker such as the social understanding tasks. Whether a differentiation into different contet domains still makes sense, is yet left open. Anyway, test takers need instructions to imagine as many and as diverse explanations for the situation (or solutions to the problem, the range of possible emotions, etc). Answers seem only possible by the use of open-ended response format relying on written language. The scoring requires a scoring key that accounts for the quanitity and diversity of the responses. Therefore, an a-priori differentiation of possible answers is valuable or must be derived from the answers of the first sample that the test is applied to. Concluding from the results of the present study and the aforementioned considerations, it would be expected that social creativity is more largely related to social understanding than to social memory or perception. The effect of written open-ended response format, however, is to be accounted or controlled for during scale construction.

Social Knowledge

Social knowledge does not belong to the core performance model of cognitive social intelligence since it is supposed to be strongly influenced by culture and the learning environment. Moreover, it is supposed to be situation-specific in such sense that only allows a judgment as correct depending on the prevalent situation. It is seen as having an impact on all remaining cognitive processes by, for example, guiding the encoding of cues into existing categories, and by influencing top-down controlled interpretation of social cues (Bless et al., 2004). The pivotal role of knowledge is also stressed by Bernieri (2001) who claims that knowledge represents the integrative component between the cognitive ability structure and the final social behavior. Bernieri's conception sees social knowledge to be the final determinant of social behavior when one knows what to do. In this case, the final behavior just can be influenced by the situation, a current intervening state or personality traits; the cognitive abilities no longer play a relevant role in such situation. In different instances, social cognitive research also claims that knowledge has an impact on the concurrent cognitive processes (see above).

Conclusively, the relevance of social knowledge is undoubted. The investigation of social knowledge represents an admittedly interesting though challenging objective for future research. Just as for the domain of social creativity, a preliminary test approach to the assessment of social knowledge shall be provided: Prior to any other steps, the knowledge domain under interest needs to be defined and delimited in terms of scope, breadth, and the classification of possible subdomains of this content area (e.g., knowledge in the domain "peer relationships"). For this objective, it appears valuable to cooperate with experts in the respective knowledge domain. The problem of expert nomination needs to be accounted for in this context, but shall not be addressed any further at this point. The establishent of a sample of critical incidents allows the later selection of different items and distracters as the foundation for the final item material.

As a next step, task material and the knowledge contents have to be sampled in accordance with the identifed structure of this content area. The sampling of adequate item material, distracters, and the identification of the "correct" solution represents the largest obstacle during the construction of a social knowledge test. One idea to account for these problems deals with the application of the Situational Judgment Test paradigm which is purported to assess tacit knowledge in a content domain. For test construction, critical incidents occurring in the kowledge domains of interest can be sampled including the true and other possible outcomes in the situation. As many as possible critical incidents should be sampled in order to account for the differentiation of the content domain. Various ways to identify the "correct" answers to the critical incidents exist, the most prominent probably is the application of an expert sample. To validate the final knowledge test, the denominated experts are supposed to show better performance than lay persons in this field.

To conclude, the final Chapters hopefully succeeded in pointing towards valuable future research directions and also provide necessary and fruitful steps for continuative test development. Although incorrect by the matter-of-fact, Goleman's (2006) denomination of social intelligence as "The new science of human relationships" is hopefully directive in such that future reserach accounts for the relevance and opportunities related to social intelligence; and for the necessary requirements to establish social intelligence as a new human ability construct.

11 References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3-27.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219-245.

Alter, U., & Muff, A. (1979). *Gewinnung von Sozialen Indikatoren für Depressives Verhalten mit Hilfe der Delphi-Methode [Extraction of Social Indicators for Depressive Behavior by the Delphi-Technique]*. Forschungsbericht des Psychologischen Instituts der Universität Zürich.

Ambady, N., LaPlante, D., & Johnson, E. (2001). Thin-slice judgements as a measure of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 89-101). Mahwah, NJ: LEA.

Amelang, M., Schwarz, G., & Wegemund, A. (1989). Soziale Intelligenz als Trait-Konstrukt und Test-Konzept bei der Analyse von Verhaltenshäufigkeiten [Social intelligence as trait-construct and measurement-concept for the analysis of act frequencies]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 10*, 37-57.

Amelang, M., & Steinmayr, R. (2006). Is there validity increment for tests of emotional intelligence explaining variance of performance criteria? *Intelligence, 34*, 459-468.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (2. erweiterte und revidierte Auflage). [Intelligence Structure Test 2000 R]*. Göttingen: Hogrefe.

Angleitner, A., Ostendorf, F., & John, O. P. (1990). Towards a taxonomy of personality descriptors in German: A psycho-lexical study. *European Journal of Personality, 4*, 89-118.

Archer, D., & Akert, R. M. (1980). The encoding of meaning: A test of three theories of social interaction. *Sociological Inquiry, 50*, 393-419.

Asendorpf, J. B. (2002). Emotionale Intelligenz nein, emotionale Kompetenzen ja [Emotional intelligence no, emotional competences yes]. *Zeitschrift für Personalpsychologie, 3*, 180-181.

Ashkanasy, N. M., & Daus, C. S. (2005). Rumors of the death of emotional intelligence in organizational behavior are vastly exaggerated. *Journal of Organizational Behavior, 26*, 441-452.

Austin, E. J. (2004). An investigation of the relationship between trait emotional intelligence and emotional task performance. *Personality and Individual Differences, 36*, 1855-1864.

Austin, E. J., & Saklofske, D. H. (2005). Far too many intelligences? On the communalities and differences between social, practical, and emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *An International Handbook of Emotional Intelligence* (pp. 107-128). Göttingen: Hogrefe.

Bänziger, T. (2005). MERT. *Multimodal Emotion Recognition Test.* Presentation, Geneva Affective Science Week. Université de Geneve, Switzerland.

Baltes, P. B., & Smith, J. (1990). Toward a psychology of wisdom and its ontogenesis. In R. J. Sternberg (Ed.), *Wisdom: Its nature, origins, and development* (pp. 87-120). New York: Cambridge University Press.

Baltes, P. B., Staudinger, U. M., Maercker, A., & Smith, J. (1995). People nominated as wise: a comparative study of wisdom-related knowledge. *Psychology and Aging, 10*, 155-166.

Bar-On, R. (1997). *The Emotional Quotient Inventory (EQ-I): Technical Manual*. Toronto: Multi-Health System.

Barchard, K. A. (2003). Does Emotional Intelligence assist in the prediction of academic success? *Educational and Psychological Measurement, Vol. 63 No. 5,* 840 – 858.

Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. *Intelligence, 13*, 263-287.

Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin Model of Intelligence Structure (BIS). *European Journal of Psychological Assessment, 30*, 977-994.

Bentler, P. M. (1992). *EQS: Structural Equations Program Manual*. Los Angeles, CA: BMDP Statistical Software.

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 3-20). Mahwah, NJ: LEA.

Bless, H., Fiedler, K., & Strack, F. (2004). *Social Cognition: How Individuals Construct Social Reality*. Philadelphia: Psychology Press.

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.

Bowman, D. B., Markham, P. M., & Roberts, R. D. (2002). Expanding the frontiers of human cognitive abilities: so much more than (plain) g! *Learning and Individual Differences, 13*, 127-158.

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.

Brody, N. (2000). History of theories and measurements of Intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 16-33). New York: Cambridge University Press.

Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry, 15*, 234-238.

Brown, L. T., & Anthony, R. G. (1990). Continuing the search for social intelligence. *Personality and Individual Differences*, *11*, 463-470.

Bronfenbrenner, U., Harding, J., & Gallwey, M. (1958). The measurement of skill in social perception. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodtbeck (Eds.), *Talent and Society* (pp. 29-111). Princeton: Van Nostrand.

Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of Social Psychology* (pp. 634-654). Cambridge, Mass.: Addison-Wesley.

Brunner, M., & Süß, H.-M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational & Psychological Measurement, 65*, 227–240.

Buck, R. (1976). A test of nonverbal receiving ability: Preliminary studies. *Human Communications Research*, 2, 162-171.

Buck, R. (1983). Nonverbal receiving ability. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal Interaction* (pp. 209-242). Beverly Hills, CA: Sage.

Burns, N. R., Bastian, V., & Nettelbeck, T. (2007). Emotional intelligence: More than personality and cognitive ability? In G. Matthews, M. Zeidner and R.D. Roberts (Eds.), *Emotional Intelligence: Knowns and Unknowns* (pp. 167 - 196). Oxford: Oxford University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Cantor, N., & Harlowe, R. E. (1994). Social intelligence and personality : flexible life-tasks pursuit. In R. J. Sternberg & P. Ruzgis (Eds.), *Personality and Intelligence* (pp. 137-168). New York: Cambridge University Press.

Cantor, N., & Kihlstrom, J. F. (1987). *Personality and Social Intelligence*. Englewood Cliffs, NJ: Prentice Hall.

Carpenter, L. J., & Merkel, W. T. (1988). The effect of three methods of observation on couples in interactional research. *The American Journal of Family Therapy, 16*, 144-157.

Carroll, J. B. (1992). Cognitive abilities: the state of the art. *Psychological Science, 3*, 266-270.

Carroll, J. B. (1993). *Human Cognitive Abilities. A survey of Factor-Analytical Studies*. New York: Cambridge University Press.

Cattell., R. B. (1950). *Personality: A Systematic Theoretical and Factual Study*. New York: McGraw-Hill.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Cattell, R. B. (1965). *The Scientific Analysis of Personality*. Harmondswort, UK: Penguin.

Cattell, R. B. (1971). *Abilities: Their Structure and Function*. Boston: Houghton Mifflin.

Cattell, R. B.. (1987). *Intelligence: Its Structure Growth and Action*. Amsterdam: North Holland.

Ciarrochi, J. V., Scott, G., Deane, F. P., & Heaven, P. C. L. (2003). Relations between social and emotional competence and mental health: a construct validation study. *Personality and Individual Differences, 35*, 1947-1963.

Chapin, F. S. (1967). *The Social Insight Test*. Palo Alto, CA: Consulting Psychologists Press.

Cline, V. B. (1964). Interpersonal perception. In B. A. Maher (Ed.), *Progress in Experimental Personality Research*. (pp. 221-284). New York: Academic Press.

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 47-65). Mahwah, NJ: LEA.

Coren, S., & Hakstian, A. R. (1992). The development and cross-validation of a self-report inventory to assess pure-tone threshold hearing sensitivity. *Journal of Speech and Hearing Research, 35*, 921-928.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five Factor Inventory Professional Manual*. Odessa, Fl: Psychological Assessment Ressources.

Costanzo, M., & Archer, D. (1993). *The Interpersonal Perception Task-15 (IPT-15)* [Videotape]. Berkeley: University of California Extension Media Center.

Cronbach, L. J. (1955). Processes affecting scores on “understanding of others” and “assumed similarity”. *Psychological Bulletin, 52*, 177-193.

Daneman, K., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Behavior, 19*, 450-466.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: in search of an elusive construct. *Journal of Personality and Social Psychology, 75*, 989-1015.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology, 10*, No. 85.

Deary, I. J., Head, B., & Egan, V. (1989). Auditory inspection time, intelligence and pitch discrimination. *Intelligence, 13*, 135-147.

Dymond, R. F. (1949). Personality and empathy. *Journal of Consulting Psychology, 14*, 343-350.

Ekman, P. (1999). Basic emotions. In. T. Dalgleish & M. J. Power, (Eds.), *Handbook of Cognition and Emotion* (pp. 45-60). New York: John Wiley.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *The Facial Action Coding System*. London: Weidenfeld & Nicolson.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech, judgments of personality and affect. *Journal of Personality and Social Psychology, 38*, 270-277.

Enzmann, D. (1996). Gestreßt, erschöpft oder ausgebrannt? Einflüsse von Arbeitssituation, Empathie und Coping auf den Burnoutprozeß [Stressed, exhausted, burned-out? Influence of the working situation, empathy, and coping on the burn-out process]. *Prävention und psychosoziale Gesundheitsforschung [Prevention and Psycho-Social Health Research]* Forschungsberichte, Vol. 3. München: Profil, Freie Universität Berlin, Fachbereich Philosophie und Sozialwissenschaften I.

Eysenck, H. J., & Eysenck, S. G. B. (1969). Personality Structure and Measurement. London: Routledge & Kegan Paul.

Fahrenberg, J., Hampel, R., & Selg, H. (2001). *Das Freiburger Persönlichkeitsinventar (FPI-R)* (7., überarbeitete und neu normierte Auflage). Göttingen: Hogrefe.

Feffer, M. H. (1959). The cognitive implications of role-taking behavior. *Journal of Personality, 27*, 152-168.

Feigenspan, J. (2005). *Computerversion des Berliner Intelligenzstruktur Tests*. Unpublished diploma thesis. University of Magdeburg, Germany.

Fidell, L. S., & Tabachnik, B. G. (2003). Preparatory data analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Comprehensive Handbook of Psychology Volume 2: Research Methods in Psychology*. New York: Wiley.

Fisseni, H.-J., & Preusser, I. (2007). *Assessment-Center. Eine Einführung in Theorie und Praxis [Assessment-Center. An Introduction in Theory and Application]*. Göttingen: Hogrefe.

Ford, M. E. (1982). Social cognition and social competence in adolescence. *Developmental Psychology, 18*, 323-340.

Ford, M. E. (1986). For all practical purposes: criteria for defining and evaluating practical intelligence. In R. J. Sternberg & R. K. Wagner (Eds.). *Practical intelligence. Nature and origin of competence in the everyday world* (pp. 183 -200). New York: Cambridge University Press.

Ford, M. E. (1992). *Human Motivation: Goals, Emotions, and Personal Agency Beliefs*. Newbury Park, CA: Sage.

Ford, M. E. (1994). A new conceptualization of social intelligence. In R.J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 974-978). NY: Macmillan Publishing Company.

Ford, M. E., & Tisak, M. S. (1983). A further search for social intelligence. *Journal of Educational Psychology, 75*, 196-206.

Forgas, J. P. (2000). *Feeling and Thinking: The Role of Affect in Social Cognition*. New York: Cambridge University Press.

Frederiksen, N., Carlson, S., & Ward, W. C. (1984). The place of social intelligence in a taxonomy of cognitive abilities. *Intelligence, 8*, 315-337.

Freudenthaler, H. H., & Neubauer, A. C. (2004). *Emotional Intelligence or Emotional Competence? The impact of the instruction on the relations of emotion-related abilities to cognitive intelligence and personality*. Presentation at the 12th European Conference on Personality, Groningen, Netherlands.

Freudenthaler, H. H., & Neubauer, A. C. (2005). Emotional intelligence: The convergent and discriminant validities of intra- and interpersonal emotional abilities. *Personality and Individual Differences, 39*, 569-579.

Freudenthaler, H. H., & Neubauer, A. C. (2007). Measuring emotional management abilities: Further evidence of the importance to distinguish between typical and maximum performance. *Personality and Individual Differences, 42*, 1561-1572.

Funder, D. C. (2001). Three trends in current research on person perception: Positivity, Realism, and Sophistication. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 319-331). Mahwah, NJ: LEA.

Funke, U., & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment, 6*, 115-123.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62*, 411-422.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Geher, G., Warner, R. M., & Brown, A. S. (2001). Predictive validity of the emotional accuracy research scale. *Intelligence, 29*, 373-388.

Gohm, C. L. (2003). Mood regulation and emotional intelligence: Individual differences. *Journal of Personality and Social Psychology, 84*, 594-607.

Goleman, D. (1995). *Emotional Intelligence – Why It Can Matter More Than IQ?*. New York: Bantam Books.

Goleman, D. (2006). *Social Intelligence: the new science of human relationships*. London: Hutchinson.

Goldberg, L. R. (1990). An alternative "Description of Personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence, 31*, 343-397.

Gough, H. G. (1968). *Manual for the Chapin Social Insight Test*. Palo Alto, CA: Consulting Psychologists Press.

Greif, S. (1987). Soziale Kompetenzen [Social competences]. In D. Frey & S. Greif (Hrsg.), *Sozialpsychologie – Ein Handbuch in Schlüsselbegriffen [Social Psychology - A Handbook of Key Concepts]* (pp. 312-320). München: Psychologie Verlags Union.

Grigorenko, E. L. (2000). Heritability and intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 53-91). New York: Cambridge University Press.

Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.

Guilford, J. P. (1977). *Way Beyond the IQ: Guide to Improving Intelligence and Creativity*. New York: McGraw-Hill.

Guilford, J. P. (1981). Higher-Order Structure-Of-Intellect Abilities. *Multivariate Behavioral Research*, 16, 411-435.

Guttmann, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence, 15*, 79-103.

Häcker, H., & Stapf, K.-H. (Eds.) (1994). Dorsch – Psychologisches Wörterbuch [Dorsch – Psychological Encyclopedia]. Bern: Huber.

Hall, J. A. (2001). The PONS Test and the Psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 143-160). Mahwah, NJ: LEA.

Hathaway S., McKinley, J., & Engel, R. (2000). *Minnesota Multiphasic Personality Inventory 2: MMPI 2*. Bern: Huber.

Hedlund, J., & Sternberg, R. J. (2000). Too many intelligences? Integrating social, emotional and practical intelligence. In R. Bar-On & J. D. A. Parker (Eds.). *Handbook of Emotional Intelligence* (pp. 136 – 167). San Francisco: Jossey-Bass.

Hendricks, M., Guilford, J. P., & Hoepfner, R. (1969). Measuring creative social intelligence. *Psychological Laboratory Reports*, Vol. 42. Los Angeles: University of Southern California.

Henry, P. J., Sternberg, R. J., & Grigorenko, E. L. (2005). Capturing successful intelligence through measures of analytic, creative and practical skills. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 295-311). Thousand Oaks: Sage Publications.

Herzmann, G., Danthiir, V., Wilhelm, O., Sommer, W., & Schacht, A. (2007). Face memory: a cognitive and psychophysiological approach to the assessment of antecedents of emotional intelligence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *Emotional Intelligence: Knowns and Unknowns* (pp. 305 - 336). Oxford: Oxford University Press.

Holliday, S. G., & Chandler, M. J. (1986). *Wisdom: Explorations in adult competence.* Basel: Karger.

Horowitz, L. M., Strauß, B., & Kordy, H. (2000). *Inventar zur Erfassung interpersonaler Probleme (IIP-D) (Deutsche Version) (2., überarbeitete und neu normierte Auflage).* Göttingen: Hogrefe.

Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 219-241). Mahwah, NJ: LEA.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic Social Cognition: Empathic Accuracy in Mixed-Sex Dyads. *Journal of Personality and Social Psychology, 59*, 730-742.

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells [Multimodal classification of intelligence performance, Experimentally controlled development of a descriptive intelligence structure model]. *Diagnostica, 28*, 195-226.

Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven [Intelligence structure research: competing models, new developments, perspectives]. *Psychologische Rundschau, 35*, 21-35.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test [Berlin Intelligence Structure Test]*. Göttingen, Germany: Hogrefe.

Jensen, A. R. (1997). *The g Factor. The Science of Mental Ability*. New York: Praeger.

Jones, K., & Day, J. D. (1997). Discriminations of two aspects of cognitive-social intelligence from academic intelligence. *Journal of Educational Psychology, 89*, 486-497.

Kang, S.-M., Day, J. D., & Meara, N. M. (2005). Social and emotional intelligence: Starting a conversation about their similarities and differences. In R. Schulze & R. D. Roberts (Eds.), *An International Handbook of Emotional Intelligence* (pp. 91-105). Göttingen: Hogrefe.

Kanning, U. P. (2002). Soziale Kompetenz - Definition, Strukturen und Prozesse [Social competence – Definition, structure, and processes]. *Zeitschrift für Psychologie, 210*, 154-163.

Keating, D. P. (1978). A search for social intelligence. *Journal of Educational Psychology, 70*, 218-223.

Kenny, D. A. (1994). *Interpersonal Perception: A Social Relations Analysis*. New York: Guilford.

Kenny, D. A., & Winquist, L. (2001). The measurement of interpersonal sensitivity: Considerations of design, components, and unit of analysis. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 265-302). Mahwah, NJ: LEA.

Kenny, D. A., West, T. V., Malloy, T. E., Albright, L. (2006). Componential Analysis of interpersonal perception data. *Personality and Social Psychology Review*, 10, 282-294.

Kessler, H., Bayerl, P., Deighton, R. M., & Traue, H.C. (2002). Facially Expressed Emotion Labeling (FEEL): PC-gestützter Test zur Emotionserkennung. *Verhaltenstherapie und Verhaltensmedizin, 23*, 297-306).

Kleinmann, M. (1997). *Assessment-Center. Stand der Forschung – Konsequenzen für die Praxis*. Göttingen: Hogrefe.

Kosmitzki, C., & John, O. P. (1993). The implicit use of explicit conceptions of social intelligence. *Personality and Individual Differences, 15*, 11-23.

Kunzmann, U., & Baltes, P.B. (2003). Beyond the traditional scope of intelligence: wisdom in action. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of Intelligence International Perspectives* (pp. 329 – 343). Washington, DC, US: American Psychological Association.

Kunzmann, U., & Baltes, P. B. (2005). The psychology of wisdom: Theoretical and empirical challenges. In R. J. Sternberg & J. Jordan (Eds.), *Handbook of wisdom: Psychological perspectives* (pp. 110-135). New York: Cambridge University Press.

Kunzmann, U., & Richter, D. (2004). Leitfaden für die Testaufgaben zur Erfassung von Empathie [Manual for tasks to assess empathy]. Jacobs Center for the Study of Lifelong Learning and Institutional Development (International University Bremen): Unpublished manuscript.

Kyllonen, P. C., & Lee, S. (2005). Assessing problem solving in context. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 11-26). Thousand Oaks: Sage Publications.

Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.) (2007). *Enhancing intelligence: Extending new constructs*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *Journal of Organizational Behavior*, *26*, 411-424.

Landy, F. J. (2006). The long, frustrating, and fruitless search for social intelligence: A cautionary tale. In K. R. Murphy (Ed.), *A Critique of Emotional Intelligence – What Are the Problems and How Can They Be Fixed?* (pp. 81-123). New Jersey: Lawrence Erlbaum Associates.

Lane, R. D., Quinlan, D. N., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The Levels of Emotional Awareness Scale: A cognitive-developmental measure of emotion. *Journal of Personality Assessment, 55*, 124-134.

Lane, R. D., Sechrest, L., Riedel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic Medicine, 58* (6), 203-210.

Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.

Lee, J.-E., Day, J. D., Meara, N. M., & Maxwell, S. (2002). Discrimination of social knowledge and its flexible application from creativity: a multitrait-multimethod approach. *Personality and Individual Differences, 32*, 913-928.

Lee, J.-E., Wong, C. T., Day, J. D., Maxwell, S. E., & Thorpe, P. (2000). Social and academic intelligences: a multitrait-multimethod study of their crystallized and fluid characteristics. *Personality and Individual Differences, 29*, 539-553.

Legree, P. J. (1995). Evidence for an oblique social intelligence factor. *Intelligence, 21*, 247-266.

Legree, P. J., Psotka, J., Tremble, T., & Bourne, D. R. (2005). Using Consensus based measurement to assess emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *An International Handbook of Emotional Intelligence* (pp. 156-179). Göttingen: Hogrefe.

MacCann, C. E. (2006). *New Approaches to Measuring Emotional Intelligence: Exploring Methodological Issues with Two New Assessment Tools*. Unpublished doctoral thesis. University of Sydney, Australia.

MacCann, C. E., Roberts, R. D., Matthews, G., & Zeidner, M. (2004). Consensus scoring and empirical option weighting of performance-based Emotional Intelligence (EI) tests. *Personality and Individual Differences, 36*, 645-662.

Marlowe, H. A., Jr. (1986). Social intelligence: evidence for multidimensionality and construct independence. *Journal of Educational Psychology, 78*, 52-58.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., Yrizarry, N., Loewinger, S., Uchida, H., Yee, A., Amo, L., & Goh, A. (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., Zeidner, M., & Roberts, R. D. (2002). *Emotional Intelligence: Science or Myth*. Cambridge, MA: MIT Press.

Matthews, G., Zeidner, M., & Roberts, R. D. (2005). Emotional intelligence – An elusive ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 79-99). Thousand Oaks: Sage Publications

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298.

Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 17*, 89-113.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence. In P. Salovey & D. Sluyter (Eds.), *Emotional Development and Emotional Intelligence: Implications for Educators* (pp. 3-31). New York: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 396–420). New York: Cambridge University Press.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitenarios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 1*, 232-242.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2002). *The Mayer, Salovey, and Caruso Emotional Intelligence Test: Technical Manual.* Toronto, ON: Multi-Health Systems.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.

McDaniel, M.A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.

McGrew, K. S. & Evans, J. J. (2004). Internal and External factorial extensions to the Cattell-Horn-Carroll (CHC) theory of cognitive abilities: a review of factor analytic research since Carroll's Seminal 1993 Treatise. *Carroll Human Cognitive Abilities (HCA) Project research Report # 2.* Evans Consulting: Institute of Applied Psychometrics.

Mehrabian, A., & Ferris, S.R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*, 248-252.

Mercer, J. R., Gomez-Palacio, M., & Padilla, E. (1986). The development of practical intelligence in cross-cultural perspective. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical Intelligence. Nature and Origins of Competence in the Everyday World* (pp. 307-337). New York: Cambridge University Press.

Moss, F. A., & Hunt, T. (1927). Are you socially intelligent? *Scientific American, 137*, 108-110.

Moss, F. A., Hunt, T., Omwake, K. T., & Woodward, L. G. (1955). *Manual for the George Washington University Series Social Intelligence Test*. Washington, D.C.: The Center for Psychological Service.

Murray, H. A. (1943). *Thematic Apperceptive Test Manual*. Cambridge, Mass.: Harvard University Press.

Neisser, V. (1976). General, academic and artificial intelligence. In L. Resnick (Ed.), *The Nature of Intelligence* (pp. 135-144). Hillsdale, N.J.: Erlbaum.

Neubauer, A. C. (1995). *Intelligenz und Geschwindigkeit der Informationsverarbeitung [Intelligence and Speed of Information Processing]*. New York: Springer.

Neubauer, A. C., & Freudenthaler, H. H. (2002). Sind emotionale Traits als Fähigkeiten messbar? [Can emotional traits be measured as abilities?]. *Zeitschrift für Personalpsychologie, 3*, 177-178.

Neubauer, A. C., & Freudenthaler, H. H. (2005). Models of Emotional Intelligence. In R. Schulze & R. D. Roberts (Eds.), *An International Handbook of Emotional Intelligence* (pp. 31-50). Göttingen: Hogrefe.

Nevo, B. (1993). In search of a correctness typology for intelligence. *New Ideas in Psychology, 11*, 391-397.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect, The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 18*, 9-35.

Nowicki, S., Jr., & 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.* Mahwah, NJ: Lawrence Erlbaum Associates.

O'Sullivan, M. (1983). Measuring individual differences. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal Interaction* (pp. 243-269). Beverly Hills, CA: Sage.

O'Sullivan, M. (2007). Trolling for trout, trawling for tuna: The methodological morass in measuring emotional intelligence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The Science of Emotional Intelligence: Knowns and Unknowns* (pp. 258 - 287). Oxford: Oxford University Press.

O'Sullivan, M., & Guilford, J. P. (1966). *Six Factor Test of Social Intelligence, manual of instructions and interpretations*. Beverly Hills, CA: Sheridan Psychological Services.

O'Sullivan, M., & Guilford, J. P. (1976). *Four Factor Tests of Social Intelligence: Manual of instructions and interpretations*. Orange, CA: Sheridan Psychological Services.

O'Sullivan, M., Guilford, J. P., & DeMille, R. (1965). The measurement of social intelligence. *Psychological Laboratory Reports*, Vol. 34. Los Angeles: University of Southern California.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: storage, processing, supervision, and coordination. *Intelligence, 31*, 167-193.

Orlik, P. (1978). Soziale Intelligenz [Social intelligence]. In K. J. Klauer (Ed.), *Handbuch der Pädagogischen Diagnostik [Handbook of Pedagogical Diagnostics]* (pp. 341-354). Düsseldorf, Germany: Schwann.

Papenbrock, J. (2005). *Geschlechtsunterschiede im Bereich der auditiven Fähigkeiten [Gender Differences in the Context of Auditory Abilities]*. Unpublished diploma thesis. Otto-von-Guericke University, Magdeburg, Germany.

Paulhus, D. L., Lysy, D. C., Yik, M. S. N. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525-554.

Pérez, J. C., Petrides, K. V., & Furnham, A. (2005). Measuring trait emotional intelligence. In R. Schulze and R. D. Roberts (Eds.), *International Handbook of Emotional Intelligence* (pp. 181-201). Cambridge, MA: Hogrefe & Huber.

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15*, 425-448.

Probst, P. (1973). Faktorenanalytische Untersuchung zum Konstrukt der „Sozialen Intelligenz“ [An Empirical Investigation of the Construct ”Social Intelligence“]. Doctoral thesis. University of Hamburg, Germany.

Probst, P. (1975). Eine empirische Untersuchung zum Konstrukt der sozialen Intelligenz [An empirical investigation of the construct social intelligence]. *Diagnostica, 21*, 24-47.

Probst, P. (1982). Empirische Untersuchung zum Konstrukt der „sozialen“ Intelligenz [Empirical investigation on the ‚social’ intelligence construct]. In K. Pawlik (Ed.), *Multivariate Persönlichkeitsforschung [Multivariate Personality Research]* (pp. 201-226). Bern: Hans Huber.

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology, 51*, 649-660.

Riggio, R. E. (1989). *Manual for the Social Skills Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Riggio, R. E., Messamer, J., & Throckmorton, B. (1991). Social and academic intelligence: conceptually distinct but overlapping constructs. *Personality and Individual Differences, 12*, 695-702.

Roberts, R. D., Schulze, R., O’Brian, K., MacCann, C. E., 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.

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion, 1*, 196-231.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnston (Eds.), *Appraisal Processes in Emotion. Theory, Methods, Research* (pp. 68 – 91). Oxford: University Press.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to Nonverbal Communication: The PONS Test*. Baltimore, MD: Johns Hopkins University Press.

Rüsseler, J., & Münte, T. F. (2001). *Readspeed*. Unpublished Test. University of Magdeburg, Germany.

Sabatelli, R. M., Buck, R., Dreyer, A (1980). Communication via facial cues in intimate dyads. *Personality and Social Psychology Bulletin, 6*, 242-247.

Sabatelli, R. M., Buck, R., Dreyer, A. (1982). Nonverbal communication accuracy in married couples: Relationship with marital complaints. Journal of *Personality and Social Psychology, 43*, 1088-1097.

Saklofske, D. H., Austin, E. J., & Minski, P. S. (2003). Factor structure and validity of a trait emotional intelligence measure. *Personality and Individual Differences, 34,* 707 – 721.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9*, 185-211.

Sander, N. (2005). *Inhibitory and executive functions in cognitive psychology: an individual differences approach examining structure and overlap with working memory capacity and intelligence*. Unpublished doctoral thesis. University of Magdeburg, Germany.

Schaie, K. W. (1996). *Intellectual Development in Adulthood: The Seattle Longitudinal Study*. Melbourne: Cambridge University Press.

Schaie, K. W. (2001). Emotional intelligence: psychometric status and developmental characteristics – Comment on Roberts, Zeidner, & Matthews (2001). *Emotion, 1*, 243-248.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York: Oxford University Press.

Scherer, K. R. (2007). Componential emotion theory can inform models of emotional competence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The Science of Emotional Intelligence: Knowns and Unknowns*. (pp. 101-126). Oxford: Oxford University Press.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion interferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76-92.

Scherer, K. R., Schorr, A., & Johnston, T. (Eds.) (2001). *Appraisal Processes in Emotion. Theory, Methods, Research*. Oxford: University Press.

Schmidt-Atzert, L. (2002). Das Mögliche machen [Doing what is possible]. *Zeitschrift für Personalpsychologie, 3*, 24-26.

Schmidt-Atzert, L., & Bühner, M. (2002, September). *Entwicklung eines Leistungstests zur emotionalen Intelligenz [Development of an Emotional Intelligence Performance Test]*. 43. Kongress der deutschen Gesellschaft für Psychologie. Humboldt-Universität Berlin.

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 2*, 8-9.

Schneider, R. J., Ackerman, P. L., & Kanfer, R. (1996). To "act wisely in human relations": exploring the dimensions of social competence. *Personality and Individual Differences, 4,* 469-481.

Schuler, H. (2002). Emotionale Intelligenz – Ein irreführender und unnötiger Begriff [Emotional intelligence – A misguiding and unnecessary Concept]. *Zeitschrift für Personalpsychologie, 3*, 138-140.

Schulze, R., Wilhelm, O., & Kyllonen, P. C. (2007). New approaches to assessing emotional intelligence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *Emotional Intelligence: Knowns and Unknowns* (pp. 199-229). Oxford: Oxford University Press.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25*, 167 – 177.

Seidel, K. K. (2007). *Social Intelligence and Auditory Intelligence – Useful Constructs?* Unpublished doctoral thesis. Otto-von-Guericke-University, Magdeburg, Germany.

Snodgrass, S. E. (2001). Correlational method for assessing interpersonal sensitivity within dyadic interactions. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity* (pp. 201-218). Mahwah, NJ: LEA.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Spearman, C. (1927). *The Abilities of Men*. London: Macmillan.

Stankov, L. (1994). Auditory abilities. In R. J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 157-162). New York Macmillan.

Stankov, L. (1999). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and Individual Differences: Process, Trait, and Content Determinants* (pp. 315-337) Washington, DC: American Psychological Association.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology*, 75, 471-490.

Staudinger, U. M., Lopez, D. F., & Baltes, P. B. (1997). The psychometric location of wisdom-related performance: intelligence, personality or more? *Personality and Social Psychology Bulletin, 23*, 1200-1214.

Staudinger, U. M., Smith, J., & Baltes, P. B. (1994). *Manual for the assessment of wisdom-related knowledge*. Berlin: Max Planck Institute for Human Development and Education.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49*, 607-627.

Sternberg, R. J. (1997). *Successful Intelligence*. New York: Plume.

Sternberg, R. J. (1998). A balance theory of wisdom. *Review of General Psychology, 2*, 347-365.

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 155–162). Norwood, New Jersey: Ablex.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 41*, 37-55.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J., Snook, S., Williams, W. M., Wagner, R. K., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.

Stricker, L. J., & Rock, D. A. (1990). Interpersonal competence, social intelligence, and general ability. *Personality and Individual Differences, 11*, 833-839.

Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen [Intelligence, Knowledge, and Problem Solving. Cognitive Determinants for Successful Performance in Computer-Simulated Problems]*. Schriftenreihe Lehr- und Forschungstexte Psychologie (Folge 5). Göttingen: Hogrefe.

Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich [The predictive validity of intelligence in the scholastic and non-scholastic domain]. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung [Perspectives of Intelligence Research]* (S. 109-136). Lengerich: Pabst Science Publisher.

Süß, H.-M. (2006). Eine Intelligenz; viele Intelligenzen? Neuere Intelligenztheorien im Widerstreit [One intelligence; many intelligences? New intelligence theories in conflict.]. In H. Wagner (Hrsg.), *Intellektuelle Hochbegabung. Aspekte der Diagnostik und Beratung [Intellectual Giftedness. Aspects in Diagnostics and Consulting.]*. Tagungsband. Bad Honnef: K. H. Bock.

Süß, H.-M., & Beauducel, A. (2005). Faceted Models of Intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 313-332). Thousand Oaks: Sage Publications.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability – and a little bit more. *Intelligence, 30*, 261-288.

Süß, H.-M., Seidel, K., & Weis, S. (2005a). *Können soziale und akademische Intelligenz separiert werden? [Can social and acedemic intelligence be separated?]*. Presentation at the 8th Conference on Differential Psychology, Personality Psychology, and Psychological Assessment. Marburg (Germany).

Süß, H.-M., Weis, S., & Seidel, K. (2005b). Soziale Kompetenzen [Social Competences]. In H. Weber & T. Rammsayer (Hrsg.), *Handbuch der Persönlichkeitspsychologie und Differentiellen Psychologie, (Reihe Handbuch der Psychologie) [Handbook of Personality Psychology and Differential Psychology]* (pp. 350-362). Göttingen: Hogrefe.

Sundberg, N. D. (1966). A method for studying sensitivity to implied meanings. *Gawein (Tijdschrift voor Psychologie)*, 15, 1-8.

Tagiuri, R. (1969). Person perception. In G. Lindzey, & E. Aronson (Eds), *The Handbook of Social Psychology* (2nd Edition, Vol. III) (pp. 395-449). Reading, Mass.: Addison-Wesley.

Thorndike, E. L. (1920). Intelligence and its use. *Harper's Magazine, 140*, 227-235.

Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin, 34*, 275–285.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38*, 406-427.

Vernon, P. A. (1993). Intelligence and neural efficiency. In D. K. Detterman (Ed.), *Current Topics in Human Intelligence* (Vol. 3, pp. 171-187). Norwood, NJ: Ablex.

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology, 4*, 42-57.

Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology, 52*, 1236 – 1247.

Wagner, R. K. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 380-395). New York: Cambridge University Press.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: the role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.

Wagner, R. K., & Sternberg, R. J. (1991). *Tacit Knowledge Inventory for Managers*. San Antonio: The Psychological Corporation Harcourt Brace & Company.

Walker, R. E., & Foley, J. M. (1973). Social intelligence: its history and measurement. *Psychological Reports, 33*, 839-864.

Weber, H., & Westmeyer, H. (2001). Die Inflation der Intelligenzen [The inflation of intelligences]. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung [Perspectives of Intelligence Research]* (pp. 251-266). Lengerich: Pabst Science Publisher.

Wechsler, D. (1981). *Wechsler adult intelligence scale, revised*. San Antonio, TX: Psychological Corporation.

Wedeck, J. (1947). The relationship between personality and ‘psychological ability.’ *British Journal of Psychology, 37*, 133-151.

Weinstein, E. A. (1969). The development of interpersonal competence. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 753-775).Chicago: Rand McNally.

Weis, S., & Süß, H.-M. (2005). Social intelligence – A review and critical discussion of measurement concepts. In R. Schulze & R. D. Roberts (Eds.), *An International Handbook of Emotional Intelligence* (pp. 203-230). Göttingen: Hogrefe.

Weis, S., & Süß, H.-M. (2007). Reviving the search for social intelligence. *Personality and Individual Differences, 42*, 3-14.

Weis, S., Seidel, K., & Süß, H.-M. (2006). Messkonzepte sozialer Intelligenz – Literaturübersicht und Ausblick [Measurement concepts of social intelligence – Review and future perspectives]. In R. Schulze, P. A. Freund, & R. D. Roberts (Hrsg.), *Emotionale Intelligenz. Ein Internationales Handbuch [Emotional Intelligence – An International Handbook]* (pp. 213-234). Göttingen: Hogrefe.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412.

Wilhelm, O. (2005). Measures of emotional intelligence: Practice and standards. In R. Schulze & R. D. Roberts (Eds.), *An International Handbook of Emotional Intelligence* (pp. 131-154). Göttingen: Hogrefe.

Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategy. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of Multivariate Experimental Psychology* (2nd ed., pp. 505-560). New York: Plenum Press.

Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem solving via Brunswik-symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *The Future of Learning and Individual Differences Research: Process, Traits, and Content* (pp. 77-104). Washington DC: American Psychological Association.

Wong, C. T., Day, J. D., Maxwell, S. E., & Meara, N. M. (1995). A multitrait-multimethod study of academic and social intelligence in college students. *Journal of Educational Psychology, 87*, 117-133.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2000). *Woodcock-Johnson III*. Itacasa, IL: Riverside.

Zeidner, M., Matthews, G., & Roberts, R. D. (2001). Slow down, you move too fast: emotional intelligence remains an ‘elusive’ intelligence. *Emotion, 1*, 265-275.

Zuckerman, M., & Larrance, D. T. (1979). Individual differences in perceived encoding and decoding abilities. In R. Rosenthal (Ed.), *Skill In Nonverbal Communication: Individual Differences* (pp. 170-195). Cambridge, MA: Oelschlaeger, Gunn & Hain, Publishers, Inc.

12 Appendix

Appendix A: Test database

Rating Test of Empathy
(Dymond, 1949; Orlik, 1978; Walker & Foley, 1973)
Test description
After a short encounter, two genuine persons A and B rate each other's personality on six dimensions on a 1-5-point scale (i.e., superior-inferior, friendly-unfriendly, leader-follower, self-confidence, selfish-unselfish, and sense of humor). The ratings are done from four different perspectives: the persons rate themselves (AA and BB), rate the other person (AB and BA), rate how the other person rates him- or herself (ABB and BAA), and rate how the other person rates oneself (ABA and BAB). A person's empathy score (person A) is determined by the differences between the two scores "ABB – BB" and "ABA – BA".
No examples
No further results

George Washington Social Intelligence Test (GWSIT)

(Moss et al., 1955; see also Thorndike & Stein, 1937)

Test descriptions

- *Judgment in Social Situations*: find possible solutions for a social problem.
- *Memory for Names and Faces*: recognize target photographs previously studied and presented later among a larger group of photographs.
- *Observation of Human Behavior*: answer questions about human functioning on a true-false basis.
- *Recognition of the Mental States Behind Words*: choose the correct mental state or emotion, among four, reflected in a given vocal statement.
- *Sense of Humor*: select the best ending of a joke.

No examples

No further results

Role Taking Test
(Feffer, 1959, as cited in Orlik, 1978 and Walker & Foley, 1973)
Test description
The test is intended to assess the cognitive developmental stage in terms of the idea of balanced decentering of the self. Test takers are presented with pictures showing stimuli like those included in the Thematic Apperception Test (Murray, 1943). They have to tell stories around the pictures based on the different roles or perspectives shown in the stimuli. Responses are judged in terms of the ability to switch between different roles.
No examples
No further results

Chapin Social Insight Test (SIT)

(Chapin, 1967; Gough, 1968)

Test description

Social insight is defined as the ability to evaluate others, to foretell what may occur in interpersonal and social situations, and the ability to rectify disturbing tensions or conflicts (Gough, 1968). The test consists of 25 items showing verbal description of different social situations. Test takers have to choose out of four alternatives the best explanation or solution to the social problem. The vignettes were drawn from case histories, literary descriptions, and published analyses of conferences and discussions as well as from prior scales for social attitudes and social adjustment.

Examples

A man bought an expensive automobile after some hesitation because it cost more than he could well afford to pay. Later, when a fried questioned him as to why he bought such an expensive car, he gave several reasons, but the one reason he did not give was:

a. His family needed to get out into the country and he bought a big car so that they could all drive together.

b. The car would save him money in the long run because it would not need the repairs that an older or cheaper car would.

c. The friend had bought a car almost as expensive although his income was not much greater.

d. He expected to receive some money from an estate by the death of a critically ill relative.

Solution "c" is correct.

No further results

Communication of Affect Receiving Ability (CARAT)

(Buck, 1976, 1983; O'Sullivan, 1983)

Test description

Genuine targets are videotaped while they are watching four types of slides (i.e., unusual, unpleasant, sexual, and scenic). The targets additionally report how pleasant they feel on a 1-9-point scale. Finally, test takers are presented 32 video scenes (of 25 different targets) without sound and are asked to judge the pleasantness on a 1-9-point scale and the type of slide that was shown. Thus, both target scoring and standards-based scoring is applied.

No examples

No further results

Profile of Nonverbal Sensitivity (PONS)

(Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979)

Test description

The PONS consists of 220 items showing one single woman simulating nonverbal responses to twenty separate social situations (e.g., comforting a lost child, helping a customer, talking about her divorce, etc.). The situations should vary on two dimensions (i.e., dominance – submission and positive – negative). The nonverbal channels are varied across the items and distinguished into a Video PONS (i.e., face vs. body vs. both) and an Audio PONS (i.e., content-filtered speech vs. randomly spliced speech). Each item presents two verbally described response alternatives in multiple choice format.

No examples

No further results

Couples Test

(Barnes & Sternberg, 1989)

Test description

Test takers are presented photographs of real and faked couples and have to decide for each of them, whether they are faked or real.

Examples

true couple

faked couple

No further results

Empathic accuracy test

(Ickes et al., 1990; Ickes, 2001)

Test description

The test is based on the dyadic interaction paradigm. 38 mixed-sex dyads were put into a short encounter and interacted with each other in an unstructured way. The interaction was recorded on videotape. Afterwards, the subjects answered questions about their thoughts and feelings during the interaction. They assigned the respective scene on the tape to each thought and feeling. They reported both the content and the valence (positive, neutral or negative) of the thoughts and feelings. For the collection of the accuracy data, the respective interaction partner viewed the videotape a second time. The investigator stopped the videotape at those times that the other subject had indicated a thought or feeling and asked the subject to infer about this thought or feeling in terms of content in valence.

The performance score for the *valence accuracy* was the percentage of matches of the dyad members' inferences about the overall emotional tone of the respective partners' entry (positive, neutral, or negative). The performance score for the *content accuracy* was the degree to which a dyad member's written description of the inferred content of his or her partner's thoughts or feelings matched the actual content that the partner had documented.

No examples

No further results

Level of Emotional Awareness Scale (LEAS)

(Lane et al., 1990; Lane, Sechrest, Riedel, Weldon, Kaszniak, & Schwartz, 1996; Ciarochi, Scott, Deane, & Heaven, 2003)

Test description

The LEAS asks the participants to describe his or her anticipated feelings and those of another person in each of 20 vignettes always involving two person (i.e., myself and another person). The degree of differentiation and integration of emotion-denoting words are rated according to 0-5 possible levels of emotional awareness.

Example *(Lane et al., 1996; p. 205)*

Example vignette:

You and your friend are in the same line of work. There is a prize given annually to the best performance of the year. The two of you work hard to win the prize. One night the winner is announced: your friend. How would you feel? How would your friend feel?

Example responses for the different levels of awareness:

0 – My friend would probably feel that the judges knew what they were doing. (nonemotion responses; “feel” describes a thought)

1 – I’d feel sick about it. It’s hard for me to say what my friend would feel. (awareness of physiological cues)

2 – I’d probably feel bad about it [...] I’m sure that my friend would be feeling really good. (relatively undifferentiated emotions; bad / good)

3 – We would both feel happy. (typical differentiated emotions)

4 – I would feel depressed. [...] I would also begrudgingly feel happy for my friend. [...] My friend would feel very gratified but [...]. (two or more Level-3-words)

5 – I’d feel disappointed [...] but glad [...]. My friend would feel happy and proud but slightly worried [...]. (scored for Level-4-responses when feelings of oneself and others were differentiated)

No further results

Tacit Knowledge Inventory for Managers (TKIM)

(Wagner & Sternberg, 1991)

Test description

The TKIM represents a Situational Judgment Test. Test takers are presented short verbal descriptions of scenarios of a situation in business settings and have to judge the effectiveness or the quality of verbally presented actions to the solution of the problem presented in the scenario. The scenarios and the possible actions were derived from experts' description of critical work-related situations. The TKIM covers three content-domains: knowledge about managing oneself, others, and tasks.

Examples

Your immediate superior has asked for your opinion on a new promotional campaign that she has developed. You think the promotional campaign is terrible, and that using it would be a big mistake. You have noticed previously that your superior does not take criticism well, and you suspect she is looking more for reassurance than for an honest opinion.

Given the present situation, rate the quality of each of the following reactions on this 1- to 7-point scale (from extremely bad to extremely good).

1. Tell her that you think the campaign is great.

2. Tell her that you like the work but have some reservations about whether it is the right campaign for this client.

Further results

Tacit Knowledge Inventories were developed for different occupational domains (i.e., business setting, university settings, military leaders, etc.). The authors claimed that practical intelligence complements *and* extends the criterion-related validity of traditional intelligence tests which are criticized to be only applied on and predictive of success in academic settings (Sternberg et al., 2000; Hedlund & Sternberg, 2000). The conceptual background of practical intelligence and tacit knowledge was frequently criticized (Gottfredson, 2003; see also Chapter 4.3.2).

Practical intelligence equaling tacit knowledge was typically assessed by just one type of measure, the Tacit Knowledge Inventory. Wagner and Sternberg (1985) and Wagner (1987) investigated expert-novice differences in specific occupational fields and found better performance of experts compared with novices in the respective domains. Academic intelligence or personality was not controlled for. However, tacit knowledge has shown to be only marginally related to academic intelligence and personality (for an overview of empirical results see Gottfredson, 2003; Hedlund & Sternberg, 2000; Henry et al., 2005; Sternberg et al., 2000). Tacit Knowledge Inventories in different occupational domains are reported to correlate partly to a substantial extent (r = .58 for academic psychology and management; r = -.06 - .36 for management and military leadership). The authors infer an underlying general ability of practical intelligence that contributed to these correlations. The criterion-related validity ranges from .05 (company prestige) to .46 (salary) (Wagner & Sternberg, 1985; Wagner, 1987). The Tacit Knowledge Inventory seems capable to assess meaningful job-related knowledge which is not equal to or explainable by academic intelligence. However, interpretations should be restricted to the assessment of knowledge without extrapolating to a generally valid intelligence construct.

Interpersonal Perception Task – 15 (IPT-15)

(Costanzo & Archer, 1993)

Test description

The IPT-15 was constructed to measure social perception and the ability to interpret expressive verbal and nonverbal behavior. Test takers watch 15 brief video scenes showing different persons in different social situations (e.g., a single woman talking to a person on the phone; two men talking to each other after a basketball game, etc.). The situations cover five postulated content dimensions, status, kinship, intimacy, competition, and deception. After each scene, test takers have to answer multiple choice questions about the social facts behind the scenes (e.g., Who is the woman talking to on the phone?, Who of the two men won the game?, etc.).

No examples

No further results

Test of wisdom-related knowledge

(Staudinger et al., 1994)

Test description

The test is based on the Berlin Wisdom Paradigm (Baltes & Smith, 1990) and intends to operationalize the 5 wisdom-related criteria (i.e., rich factual and procedural knowledge, life-span contextualism, value relativism, and uncertainty). Test takers are presented short situation descriptions that cover two thematic domains, life planning and life review. They are asked to think aloud about the presented situation or problem. Their answers are judged by trained raters according to how they match the five criteria.

Examples (Staudinger et al., 1994; p. 15)

Situation description (Target: Older female adult in a work-family-related problem in life planning)

Joyce, a 60-year-old widow, recently completed a degree in business management and opened her own business. She has been looking forward to this new challenge. She has just heard that her son has been left with two small children to care for.

Joyce is considering the following options: She can plan to give up her business and live with her son, or she can plan to arrange for financial assistance for her son to cover child-care costs.

What should Joyce do and consider in making her plans? What additional information is needed?

Further results

The test to assess wisdom-related knowledge was applied in a nomination study that investigated the validity of the Berlin Wisdom Paradigm (Baltes, Staudinger, Maercker, & Smith, 1995). People nominated as wise (according to laypersons' conception of wisdom) were compared with age-equivalent clinical psychologists, an age-equivalent comparison group and a younger comparison group (both from non-human-service occupations). Results indicated an advantage of both, the nominated group and that of clinical psychologists. The authors concluded that age is not the critical factor that determines wisdom-related performance. Moreover, important factors appeared to be general experiences, professional training and practice, and motivational preferences (Kunzmann & Baltes, 2005).

Staudinger, Lopez, and Baltes (1997) investigated the overlap of wisdom with variables of the domain of intelligence, personality, and the personality-intelligence interface (i.e., social intelligence, creativity, and cognitive styles). Social intelligence was operationalized by two self-report inventories. Results showed that intelligence, personality, and measures that straddled this interface could explain 40 % of the variance in wisdom-related performance whereby the interface measures (i.e., five measures of cognitive style and creativity) explained the largest amount of unique variance.

Situational Judgment Test of Social Intelligence

(Legree, 1995)

Test description

The scales for dinner-related knowledge requires participants to rate the relative appropriateness of 20 actions. The scale for knowledge of alcohol abuse requires participants to rate the extent to which 20 statements suggest alcohol abuse.

Examples *(Legree, 1995; pp. 265)*

Dinner-related knowledge

Assume that you are married and have two children. You and your family have been invited to attend a dinner at the home of your supervisor. In the office, you usually interact with your supervisor in a friendly and nonchalant manner, and you have known him for several years. At this dinner, you will meet his family. You want to make a good impression at the dinner because you are being considered for a promotion.

Using the rating scale (extremely inappropriate (1) – neither appropriate nor inappropriate (6) – extremely appropriate (11)), estimate the effectiveness of the following actions with respect to portraying as competent and promotable.

1. Discuss the weather
2. Describing your dislike of a pet that is nuisance to your neighborhood.
3. Dominating the conversation in order to appear confident and self-assured to your supervisor.
4. Using your fingers to eat a piece of food.
5. Complimenting the hosts on the quality of food.
6. Flattering the hosts on their "superb taste in home decoration".
7. Discussing sports or artistic issues.

Knowledge of indicators of alcohol abuse

This section provides short descriptions of the habits and actions of 20 individuals. Your task is to rate the extent to which each of the descriptions is more or less consistent with the expectation that the individual may or may not abuse alcohol (extremely unlikely (1) – neither unlikely nor likely (6) – extremely likely (11)).

1. Nancy is frequently late for work, and often calls in sick on Mondays.
2. Sue is a moderately heavy smoker.
3. At a business lunch held at Mary's favorite lunch spot, the waitress asked Mary "Ma'am, will you have your usual drink today?" Mary quickly responded, "I'll have coffee today."
4. Ever since his grandfather died three month ago, James has been in state of deep depression.
5. One can occasionally smell alcohol on Liz's breath after lunch and she frequently shuts her office door for several hours after returning from lunch.
6. Over the past 18 month, Thomas has been involved in three automobile accidents.

No further results

Emotional Accuracy Research Scale (EARS)

(Mayer & Geher, 1996; Geher et al., 2001)

Test description

Participants are given three written descriptions of situations (compose one vignette) that contribute to the present mood of a target person. After reading, they are asked to select the best adjective out of pairs of adjectives that best describes the present feelings of the target. The test consists of eight vignettes with twelve items each.

Examples *(Mayer & Geher, 1996; p. 98)*

Written by a 20-year-old woman (one of three descriptions of one vignette)

“My roommate has been kind of blowing off her boyfriend. She told him she did not want to see him until Spring Break. He is hurt because he thinks she does not like him anymore, and he wants to come up here to see her this weekend. I have been gone almost every weekend since school started, giving her plenty of opportunities to have im up here while I am gone, and now I’m finally getting to stay here for the weekend and he might be coming up. (Why can’t she go visit him instead?!?!)”

For each of the twelve pairs below, choose the word or phrase within that pair which best describes the reported feelings of the person who wrote the above passage across all the situations she described (correct answers marked with an asterisk).

“1. Be by myself* – Kick something; 2. Stomping feet – Alone*; 3. Pretend everything is ok* – Threaten a fight; 4. Angry for someone else – Help a friend*; 5. Evade feeling* – Defiant; 6. Sharing another’s anger* – Threatened with death; 7. Hostile – Unhappy with another*; 8. Fearful – Apart from another*; 9. Cheated* – My teeth clenched; 10. Withdraw – Scared for someone else*; 11. Attacked – Isolate myself*; 12. Mad* – Delighted.”

No further results

Japanese and Caucasian Brief Affect Recognition Test (JACBART)

(Matsumoto, LeRoux, Wilson-Cohn, Raroque, Kooken, Ekman et al., 2000)

Test description

The test was constructed to measure emotion recognition ability. Its scope and objectives reach beyond the, at that time, published tests of Ekman and Friesen (Ekman & Friesen 1975, cited in Matsumoto et al., 2000). The test varies systematically the ethnic group and the gender across the seven universal emotions (i.e., anger, contempt, disgust, fear, joy, sadness, and surprise). The facial stimuli were scored by the use of the Facial Action Coding System in order to verify the same facial muscles configurations associated with the emotions. Each facial emotion expression is included in a one-second-video-scene with a neutral expression at the beginning and in the end so that no after-image of the expression is possible. Furthermore, the duration of the emotion expression was varied (1/15 sec., 2/15 sec., and 2/5 sec.). After the presentation of one item, test takers have to indicate how much each of the seven emotions were present in the expressions (on a 0-8-point scale from not at all to a lot). Group consensus scoring was applied for this type of task. In a forced-choice format, test takers have to decide which of the seven emotions is pictured. This task is scored according to the standard rule for the universal emotions.

No examples

No further results

Diagnostic analysis of nonverbal accuracy scale (DANVA)

(Nowicki & Duke, 1994, 2001)

Test description

The test consists of several subscales, both for children and adults. The two most interesting subscales are those based on facial and vocal stimuli of emotion expressions. The stimuli were selected to cover a wide range of difficulty by including emotion expressions of varying intensity. Task material was posed by actors and includes stimuli from members of different racial groups.

The test is intended to measure the receptivity of four basic emotions (i.e., happiness, sadness, anger, and fear).

In both subscales, the test takers are presented each stimulus separately and they have to indicate which of the four above-mentioned emotions is expressed in the face or in the voice. Answers are scored in terms of correct vs. incorrect.

No examples

No further results

Vocal Emotion Recognition Test (Vocal-I)

(Scherer et al., 2001)

Test description

Test takers have to identify emotions in meaningless sentences spoken by male and female actors. The emotions that have to be identified are joy, sadness, fear, anger, and neutral. Both, the Vocal-I and the MERT are based on the GVEESS, the Geneva Vocal Emotion Expression Stimulus Set. The GVEESS includes 224 vocal emotion portrayals of 14 different emotions (i.e., hot anger, cold anger, panic fear, anxiety, despair, sadness, elation, happiness, interest, boredom, shame, pride, disgust, and contempt).

No Examples

No Further results

Facially Expressed Emotion Labeling (FEEL)

(Kessler, Bayerl, Deighton, & Traue, 2002)

Test description

Participants have to decide which emotion is present in facial expression of the six basic emotions (i.e., anger, sadness, disgust, fear, surprise, happiness) relying on multiple-choice response format. Answers are scored in terms of correctness. The stimuli were judged as unambiguous by the use of the Facial Action Coding System and stem from the database of Ekman, Friesen, and Hager (2002).

No examples

No further results

Verbal Social-Cognitive Flexibility (SCF-V)

(Lee et al., 2002)

Test description

The test takers are presented three vignettes. Each contains a verbal descriptions of an ambiguous social situation which involves some novelty. Participants are given 3 minutes to write down all possible interpretations of each episode. Answers are scored according to a previously developed coding scheme in terms of fluency and flexibility. The fluency score represented the number of relevant interpretations across the three episodes. The flexibility score represented the number of hits of different categories that had been generated in advance (e.g., Category I: personal characteristics, states, traits, activities; Category II: cognition, perception, thought, etc.). The fluency and the flexibility score were summed up for the final score.

Example (Lee et al., 2002, p. 918)

"In one episode, two males who routinely play racquetball together are described. One is a lawyer who encourages the other (a college student) to apply for law school. When the college student asks for assistance in studying for the law school, however, the lawyer appears uninterested in helping and later stops playing racquetball with the college student."

No further results

Pictorial Social-Cognitive Flexibility (SCF-V)

(Lee et al., 2002)

Test description

The test takers are presented three one-minute video clips each portraying male and female actors in ambiguous and novel social situations. Participants were given 3 minutes to write down all possible interpretations of each episode. Answers are scored according to a previously developed coding scheme in terms of fluency and flexibility. The fluency score represented the number of relevant interpretations across the three episodes. The flexibility score represented the number of hits of different categories that had been generated in advance (e.g., Category I: personal characteristics, states, traits, activities; Category II: cognition, perception, thought, etc.). The fluency and the flexibility score were summed up for the final score.

Example (Lee et al., 2002, p. 919)

"In one video, a man, dressed in a tuxedo, walks from his car to the front door of a house where he knocks on the door. A woman wearing jeans and a sweatshirt answers the door, walks with the man to his car, gets into the car with the man and they drive away."

No further results

Test of Emotional Intelligence (TEMINT)

(Schmidt-Atzert & Bühner, 2002, Amelang & Steinmayr, 2006)

Test description

Test takers are presented 12 situation descriptions and have to rate the presence of 10 emotions of the targets in the situations. The situations were generated by genuine persons reporting about meaningful life events and their feelings in that situation. The emotions rated were, aversion, anger, fear, unease, sadness, guilt, happiness, pride, affection, and surprise. The ratings were done on a 1-3-point scale from not at all present to strongly present. The test is scored in terms of the simple difference between judge and target answers.

Examples

Female student, 24 years old: "I have failed in an important exam and had to repeat it."

Transpose yourself into the role of the female student. How strong were her emotions?

Emotions rated: aversion, anger, fear, ...

Further results

Schmidt-Atzert and Bühner (2002, see also Amelang & Steinmayr, 2006) report only moderate correlations with personality and intelligence and found substantial correlations with school grades when controlling for intelligence (r = .24 / .42, for math and german language, respectively). However, Amelang and Steinmayr (2006) could not detect incremental validity of the TEMINT for the prediction of school grades, social status, or educational level over and above academic intelligence and conscientiousness. The academic intelligence test applied was a multidimensional measure including verbal, figural and numeric task material and separate operational ability domains (i.e., reasoning, speed of information processing, and memory). Thus, the contradictory findings could be attributed to this.

Facial Emotion Inspection Time Task

(Austin, 2004)

Test description

Respective to a choice reaction time paradigm, participants have to decide whether a shortly presented face shows an emotionally laden expression (happy and sad for the two test subscales) or a neutral expression. Presentation times were systematically varied (from 17 to 350 ms). The stimuli stem from the database of Ekman and Friesen (1976; as cited in Austin, 2004) to assure unambiguousness. Both, male and female faces were used of two different targets each.

No examples

No further results

Test for the Assessment of Empathy

(Kunzmann & Richter, 2004)

Test description

The test consists of 8 video scenes presenting target persons telling about positive or negative feelings (happiness or sadness; 4 persons for each feeling) they have had while watching an emotion-eliciting video. Test takers have to indicate how much one of a list of 20 emotions is present in one video scene. Performance is judged in terms of the correlation of the judges' and the targets' answers (target scoring) and in terms of the correlation of the targets with a trained observer who had also rated the presence of emotions (expert consensus scoring). The list of emotions contained the following elements: fearful, happy, indifferent, sad, angry, joyful, hostile, affected, amused, burdened, empathic, depressed, contemptuous, interested, saddened, worried, delighted, abject, excited, aggrieved. The targets were selected in a complex procedure that accounted for the concordance between their reported emotions, the thematic of the video they had to watch, and the judgment of the trained observer. Furthermore, they should show homogeneous personality profiles.

No examples

No further results

Multimodel Emotion Recognition Test (MERT)

(Bänziger, 2005)

Test description

The MERT like the Vocal-I is also based on the GVEESS with a database 224 vocal emotion portrayals (see the description of the Vocal-I). In a facetted design, each of 10 acted emotions is presented by four different types of task material (i.e., audio, video, audio+video, pictures) and three sentences.

No examples

No further results

Test of Emotional Abilities

(Freudenthaler & Neubauer, 2005)

Test description

According to Freudenthaler and Neubauer (2005), the test idea is based on a criticism on existing operationalization of emotional abilities in the MSCEIT. The instruction to these tasks (see Branch IV operationalizations of the MSCEIT) suggest to measure emotional knowledge about the intended constructs (i.e., knowledge about the adequate behavior for emotion regulation) instead of the actual effectiveness of emotion regulation abilities (i.e., the typical performance).

Subjects are presented emotionally laden scenarios and have to choose out of four alternatives which behavior would best describe their actual behavior in the given situation (typical performance). Answers are scored in terms of degree of correctness. In order to determine correctness, the response alternatives were rated by a panel of 10 experts for adequacy or effectiveness on a 1-4-point-scale (most adequate to least adequate).

Examples (Freudenthaler & Neubauer, 2004)

Intrapersonal emotional abilities

A close friend has told you that he doesn‘t trust and understand you. You are very sad about it. Response alternatives: a) I talk with him about it in order to find out the reason. b) I analyze the communication problem. c) I try to get his confidence back. d) I try to accept it.

Interpersonal emotional abilities

A good friend was recently abandoned by his/her partner. He/she severely suffers from it. Response alternatives: a) I suggest to go out more often together then. b) I refer to the negative attributes of the ex-partner and emphasize the advantages of being a single. c) I tell him/her that I will be available for him/her if he/she would like it. d) I assure him/her that he/she will find a new partner soon.

Further results

Freudenthaler & Neubauer (2005, 2007) could replicate results about the validity of their test. Applying the typical performance instruction resulted in larger correlations with personality traits and self-reported emotional intelligence, and lower correlations with academic intelligence tests. Applying the traditional maximum performance instruction to the same test, the reverse correlation pattern was found.

However, the authors could not provide evidence about the underlying requirements of their test beyond the dispositional tendencies assessed by personality trait inventories. Furthermore, the test could not yet prove incremental validity in the prediction of relevant external criteria.

Situational Test of Emotional Understanding (STEU)

(MacCann, 2006)

Test description

The test development was based on the taxonomic principles of the appraisal theory of the structure of emotions (Roseman, 2001). In short, the taxonomy postulates that the appraisal of emotions depends on the surrounding circumstances. These circumstances are classified according to seven dimensions (e.g., "relief is caused by appraisals of circumstance-caused, certainty, motive consistency, and aversive stimuli", p. 45). Item construction was based on this taxonomy and thus allows standards-based scoring.

Test contains 42 multiple-choice items. For each of 14 emotions, three structurally equivalent items were given: one of work context, one of private context, and one that described the abstract features of the situation and was thus context-less.

Examples (MacCann, 2006, p. 93)

"Xavier completes a difficult task on time and under budget. Xavier is most likely to feel?

(a) surprise, (b) pride, (c) relief, (d) hope, (e) joy."

Further results

(see Chapter 5.2.4 for an overview of analysis assessing the implications of different response formats)

Situational Test of Emotional Management (STEM)

(MacCann, 2006)

Test description

This test is based on the Situational Judgment Test approach and followed the construction rational suggested by McDaniel and Nguyen (2001). Item contents were derived from semi-structured interviews. They were selected according to content-domain of life and the type of emotion. Response alternatives were generated in a separate step by asking people about their actual and ideal responses. Finally, experts (i.e., counselors, emotional intelligence researchers, life coaches) were asked to respond to the items both in multiple coice format (13 experts) and ratings-based format (6 experts).

The final test contains two forms A and B (multiple-choice and ratings-based, respectively). Subjects have to identify the most effective solution / the degree of effectiveness for the solution of an emotional problem.

Examples (MacCann, 2006, p. 94)

multiple choice:

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

a. Nothing, it will sort itself out soon enough

b. Tell his family he feels left out.

c. Spend time listening and getting involved again.

d. Reflect that relationships can change over time.”

ratings-based:

“Clayton has been overseas for a long time and returns to visit his family. So much has changed that Clayton feels left out. How effective are each of the following actions for Clayton? [on 6-point scale from *not at all effective* to *extremely effective*).

(a) Nothing, it will sort itself out soon enough

(b) Tell his family he feels left out

(c) Spend time listening and getting involved again

(d) Reflect that relationships can change over time”

Further results

(see Chapter 5.2.4 for an overview of analysis assessing the implications of different response formats)

Appendix B: Letter of Information to Targets

Sehr geehrte/r Frau / Herr,

zunächst möchten wir uns recht herzlich bei Ihnen bedanken, dass Sie uns und der Wissenschaft behilflich sein möchten. Damit Sie sich endgültig dafür entscheiden können, uns bei der so genannten Itemerstellung zu helfen, möchten wir Sie im Folgenden umfassend über den geschätzten Aufwand und Ablauf und über generelle Voraussetzungen informieren. Dieses Informationsschreiben sollte Ihnen eine Einschätzung darüber ermöglichen, was genau auf Sie und gegebenenfalls auf Dritte zukommt und Sie darüber informieren, was mit dem von uns angefertigten Bild- und Tonmaterial geschieht.

Über uns

Wir sind Mitarbeiter in einem Forschungsprojekt an der Universität Magdeburg. Unser Projekt beschäftigt sich mit der Erfassung sozialer Intelligenzleistungen. Dafür möchten wir vollkommen neue bzw. neuartige Testinstrumente erstellen, die geeignet sind, sozial intelligente Leistungen zu erfassen. Unserer Auffassung nach besteht soziale Intelligenz zu einem großen Teil aus der Fähigkeit, andere Personen in unterschiedlichen Situationen gut einschätzen zu können, d.h. ihre Gefühle und Gedanken identifizieren, verstehen und ihr Verhalten interpretieren zu können. Bisherige Testinstrumente versuchen, diese Fähigkeit auf Basis völlig dekontextualisierter Informationen zu erfassen. Unserer Meinung nach entspricht dies kaum den realen Anforderungen, die an jeden sozial agierenden Menschen gestellt werden. Die mangelnden Forschungserfolge zur Identifizierung sozial intelligenter Leistungen führen wir zum Teil auf diese in unseren Augen veraltete Testidee zurück. Stattdessen ziehen wir eine Reihe von Kontextinformationen hinzu, um eine möglichst gute Einschätzung einer Person abgeben zu können. Auf Grundlage dieser Überlegungen entstand eine Testidee, die in der Frage des Kontextes von bisherigen Ansätzen abweicht. Um diese Testidee zu verwirklichen, brauchen wir Sie.

Ablauf

Wir, d.h. zwei Mitarbeiter der Universität Magdeburg, möchten Sie gerne über einen Zeitraum von einem bis drei Tagen begleiten, um in unterschiedlichen Situationen Bild-, Ton- und Videoaufnahmen von Ihnen und den mit Ihnen interagierenden Personen zu machen. Dabei sind wir sowohl an beruflichen als auch an privaten Situationen interessiert. Sie können jeweils im Voraus bestimmen, wann und in welchen Situationen Sie uns erlauben möchten, Sie zu begleiten. Eine Situationsauswahl sollte in jedem Fall vorab mit uns getroffen werden, da wir insbesondere an Interaktionssituationen interessiert sind und diese sowohl im privaten als auch im beruflichen Umfeld abbilden möchten. Denjenigen Personen, von denen Sie schon sicher wissen, dass Sie Ihnen in dieser Zeit begegnen werden, können im Vorab dieselben Informationen zugehen. Selbstverständlich müssen wir auch von diesen Personen eine Einverständniserklärung einholen.
Wir werden selbstverständlich versuchen, so unsichtbar wie möglich zu sein und Sie bitten, so wenig wie möglich Notiz von uns zu nehmen. Uns kommt es bei den Aufnahmen nicht auf ein bestimmtes Verhalten an, sondern wir sind vor allem daran interessiert, ganz natürliches Verhalten in Bild und Ton festzuhalten. Zusätzlich zu den Bild- und Tonmaterialien sollen

auch von Ihnen geschriebene Informationen gesammelt werden. Dazu zählen wir vor allem von Ihnen verfasste E-Mails, Kurznotizen, Briefe, etc. Selbstverständlich dürfen Sie auch hier darüber entscheiden, welche dieser Materialien Sie uns zukommen lassen wollen.
Während der Aufnahmen bzw. in passenden Unterbrechungen werden wir Sie über Ihre Gedanken- und Gefühlswelt befragen, die sich in den Bildern, in dem Gesprochenen und Geschriebenen widerspiegelt. Dazu werden wir Sie zu verschiedenen Zeitpunkten bitten, kurze Fragebögen auszufüllen. Diese Informationen, die Sie uns geben, werden später als Grundlage für die Testerstellung dienen. Der eigentliche Test wird später darin bestehen, dass die jeweiligen Testpersonen, denen die gesammelten Bild-, Ton- und Textmaterialien gezeigt werden, an bestimmten Stellen eine Einschätzung über die Gedanken- und Gefühlswelt der gesehenen Personen abgeben sollen. Diese Einschätzung wird mit den von den jeweiligen Personen selbst gemachten Angaben verglichen. Aus dem Grad an Übereinstimmung wird sich dann die Qualität der Leistung ergeben.
Nach Abschluss der Aufzeichnungen erhalten Sie die Möglichkeit, das gesammelte Material zu sichten und gegebenenfalls bestimmte Szenen von der weiteren Verarbeitung auszuschließen. Bei dieser abschließenden Sichtung des Materials werden wir Ihnen gegebenenfalls zu einzelnen Szenen weitere Fragen bezüglich Ihrer Gedanken und Gefühle stellen, etwa an welchen Hinweisreizen in Bild und Ton Sie Ihre eigenen Gefühle und Gedanken in der jeweiligen Szene identifizieren würden. Zuletzt werden wir Sie darum bitten, einen Fragebogen auszufüllen, der Bereiche Ihrer Persönlichkeit erfasst. Wiederum wird es hier Aufgabe der späteren Testperson sein, Ihre Persönlichkeit in diesen Bereichen möglichst zutreffend einzuschätzen.

Weiterverarbeitung

Nachdem Sie das Material abschließend gesichtet und gegebenenfalls einzelne Szenen ausgeschlossen haben, werden wir das Material weiterverarbeiten. Diese Weiterverarbeitung wird vor allem darin bestehen, das Material zu kürzen. Für den finalen Test benötigen wir ein „Szenario“ von 10 bis 15 Minuten Dauer, das die relevanten und interessierenden Informationen möglichst komprimiert aber realistisch darstellt. Beachten Sie, dass wir beim Zusammenschnitt des Materials dieses möglichst authentisch lassen wollen. Auch wir sind nicht daran interessiert, realitätsferne Informationen darzustellen, da in der Realitätsnähe der Darstellung der eigentliche Sinn der Testidee liegt.
Ziel der Nachbereitung ist es, ein Szenario zu erstellen, in dem relevante Hintergrundinformationen über Sie dargestellt werden und auf dessen Basis die Testpersonen versuchen sollen, Sie so gut wie möglich „kennen zu lernen“ und daraufhin einzuschätzen. Ob das Material dann tatsächlich in den Test integriert wird, entscheiden wir nach dem Zusammenschnitt und nach diversen Vortestungen. Bei diesen Vortestungen wird überprüft, ob die Aufgaben dazu geeignet sind, sozial intelligente Leistungen zu messen. Dabei kommt es vor allem darauf an, wie schwer es den Testpersonen fällt, die richtige Einschätzung zu treffen und ob es dabei besser oder schlechter geeignete Testpersonen gibt.
Nur für den Fall, dass das Szenario die genannten Anforderungen erfüllt, wird es in der Hauptuntersuchung eingesetzt. In der Hauptuntersuchung wird wiederum geprüft, ob wir mit den Tests (dazu gehören dann auch weitere andersartige Aufgaben, die mit der hier beschriebenen Aufgabe nichts zu tun haben) überhaupt soziale Intelligenz erfassen konnten. Erst wenn alle diese Voraussetzungen erfüllt sind, wird darüber entschieden, ob der Test in der Endform publiziert wird. Eine Testpublikation enthielte dann alle Bild-, Ton- und Textmaterialien, da die letztendlichen Testkäufer in der Lage sein müssen, die Antworten ihrer Testpersonen auf dieser Basis auszuwerten. Verlegt würde der Test von einem Fachverlag für psychologische Testverfahren. Es würde sich also um die wissenschaftliche

Publikation eines Verfahrens handeln, dass von Fachleuten erworben und eingesetzt werden soll.
Wie sie sehen, ist es zum jetzigen Zeitpunkt noch ungewiss, ob die Hürden zu einer erfolgreichen Testerstellung überwunden werden können. Sollte dies jedoch der Fall sein, werden wir den Test publizieren.

Rechtliche Aspekte / Datenschutz

Für jede Form der Aufzeichnung (in Bild, Ton und Schrift), die wir von Ihrer Person oder von Dritten sammeln und später den Testpersonen präsentieren, gilt, dass sämtliche personenbezogenen Informationen (Namen, Adressen, Arbeitsplatz, u. ä.) eliminiert werden. Damit werden keinerlei Daten über Sie, Ihre Familie oder Ihre Arbeit an Dritte weitergegeben, so dass der Datenschutz vollständig gewahrt wird. Wir sichern Ihnen außerdem zu, dass wir die Aufzeichnungen und die darin enthaltenen Informationen nicht verzerrt darstellen werden.

Wir brauchen von Ihnen zu zwei Aspekten Ihr Einverständnis:

1. Sie erklären sich damit einverstanden, dass wir Bild-, Video- und Tonaufnahmen von Ihnen anfertigen.
2. Sie erklären sich damit einverstanden, dass wir das entstandene und von Ihnen gesichtete Material zu wissenschaftlichen Zwecken nutzen und weiterverarbeiten.

Unter der Nutzung und Weiterverarbeitung zu wissenschaftlichen Zwecken verstehen wir:

- Auswahl und Schnitt des Materials zur Erstellung eines 10-15 minütigen Szenarios, das Sie in Ihrem privaten und beruflichen Umfeld darstellt
- Darbietung dieses Szenarios im Rahmen von wissenschaftlichen Untersuchungen als Grundlage für Leistungstests zur Erfassung von sozialer Intelligenz
- Publikation der Materialien im Rahmen von wissenschaftlichen Arbeiten (d.h. Präsentationen auf Fachtagungen, in Fachzeitschriften und im Rahmen von Testpublikationen)
- Nur die Testpublikation betreffend: Ihre schriftlichen Antworten in den Fragebögen zur Erfassung Ihrer Gefühls- und Gedankenwelt und Ihrer Persönlichkeit werden ausschließlich im Rahmen der Testpublikation veröffentlicht, da sie den Testkäufern zur eigenständigen Auswertung der Tests dienen muss. Diese Informationen werden den späteren Testpersonen nicht zur Verfügung gestellt.

Ihre Einverständniserklärung umfasst den Verzicht auf eine Beteiligung an etwaigen Erträgen aus Testverkäufen.

Aus urheberrechtlichen Gründen können wir Ihnen das angefertigte Bild- und Tonmaterial leider nicht aushändigen.

Mit unserer Unterschrift unter dieses Schreiben sichern wir Ihnen zu, dass wir die datenschutzrechtlichen Vorgaben erfüllen und dass wir die von Ihnen aufgenommen und gesichteten Materialien ausschließlich im Rahmen der oben aufgeführten wissenschaftlichen Zwecke nutzen werden.

Zusammenfassung

Hier noch einmal die entscheidenden Informationen im Überblick:

- Bedenken Sie den Aufwand: Je nach der Verfügbarkeit von „interessanten“ Szenen werden wir ein bis drei Tage lang zu bestimmten Tageszeiten Ihre Begleiter sein.
- Bedenken Sie, dass nicht nur Sie, sondern auch Ihnen nahestehende Dritte betroffen sein können.
- Bedenken Sie, dass Sie nach der abschließenden Sichtung des Materials keinen Einfluss auf die Weiterverarbeitung mehr haben. Allerdings sichern wir Ihnen zu, dass die Weiterverarbeitung sich auf die oben genannten Schritte beschränkt.
- Bedenken Sie, dass wir Ihnen das Bild- und Tonmaterial aus urheberrechtlichen Gründen im Falle einer Testpublikation nicht aushändigen können.
- Bedenken Sie, dass die Möglichkeit besteht, dass das Material zu einem Test verarbeitet wird, der u.U. als Leistungstest zu sozialer Intelligenz publiziert wird.

Falls dieses Schreiben noch Fragen offen lässt, wenden Sie sich bitte an uns. Wir sind gerne bereit, Unklarheiten zu beseitigen. Lesen Sie bitte auch die Einverständniserklärung nochmals gründlich durch. Sie fasst die wesentlichen Punkte nochmals zusammen.

Wir würden uns freuen, wenn Sie uns unterstützen würden. Wenn Sie sich dazu entschließen, unterschreiben Sie bitte die Einverständniserklärung in zweifacher Ausführung, ein Formular und dieses Schreiben verbleiben bei Ihnen.

Unten aufgeführt finden Sie die Namen Ihrer Ansprechpartner und die Kontaktinformationen, unter denen Sie uns erreichen können.

Wir hoffen auf eine positive Antwort Ihrerseits, auf gute Zusammenarbeit und verbleiben mit freundlichen Grüßen,

(Heinz-Martin Süß) (Kristin Seidel) (Susanne Weis)

Postadresse
Institut f. Psychologie (IPSY)
Otto-von-Guericke-Universität Magdeburg
Postfach 4120
39016 Magdeburg

Telefon

0391/67-18468 0391/67-18486 0391/67-18486

Telefon Sekretariat:

0391/67-18464

email:
heinz-martin.suess@gse-w.uni-magdeburg.de
kristin.seidel@gse-w.uni-magdeburg.de
susanne.weis@gse-w.uni-magdeburg.de

Appendix C: Course of Testing in Study 1 Including Planned Task Durations

Study 1 – Testing day 1

No.	Test	Planned Duration / min.
1	General instructions / Biographical Questionnaire	5
2	SRT	10
3	SPv1	15
4	Recognition of Repeated Tones	10
5	SMa1 Part 1	15
6	Sound Recognition	15
7	Chord-Decomposition	15
8	Break	10
9	BIS Part 1	15
10	SMf1+2	15
11	Scenario 1: Matthias (SU_MM)	15
12	Questionnaire: MF – MMPI-2	10
13	Questionnaire: Hearing Screening Inventory	5
14	Break	10
15	Mouse Task	10
16	Tonal Analogies	10
17	SPp1	15
18	SMa1 Part 2	15
19	Dissected Sentences	15
20	Break	10
21	BIS Part 2	15
22	Scenario 2: Katharina (SU_KL)	15
23	Questionnaire: Social Behavior	15
	Sum	285

Study 1 – Testing day 2

No.	Test	Planned Duration / min.
1	SPa1 Part 1	15
2	NEO-FFI	10
3	Tonal Figures	15
4	Rhythm Reproduction	15
5	Recognition of Repeated Voices	10
6	AIT-P	25
7	Break	10
8	SMp1 / SMp 2	15
9	BIS Part 3	15
10	Scenario 3: Renate (SU_RF)	15
11	SMv 1+2	20
12	Break	10
13	SPf1	15
14	Masked Words	15
15	Questionnaire: Computer Experience	5
16	Tonal Series	15
17	Audiobook	10
18	SPa1 Part 2	15
19	Break	10
20	Readspeed	5
21	BIS Part 4	15
22	BIS Part 5	15
23	Scenario 4: Christoph (SU_CP)	15
	Sum	305

BIS-Tasks (selection and order, 5 test parts)

Part 1	**cell assignment**	**IT**	**WT**	∑
VS	warming up	0:40	1:30	2:10
SI	SN	0.20	0.50	1:10
FA	RF	0:30	3:30	4:00
WM	MV	0:55	1:30	2:25
RD	RN	0:40	3:20	4:00
ZZ	MN	1:15	0:50	1:05
∑				15:50
Part 2	**cell assignment**	**IT**	**WT**	∑
SV	RV	1:30	1:30	3:00
WE	MF	0:50	0:40	1:30
WS	RV	0:30	1:00	1:30
OE	SF	0:30	0:30	1:00
TL	RN	0:20	5.00	5:20
OG	MF	1:55	1:40	3:35
∑				15:55
Part 3	**cell assignment**	**IT**	**WT**	∑
ZN	RN	0:50	3:50	4:40
KW	SV	0:15	0:30	0:45
AN	RF	1:00	1:45	2:45
XG	SN	0:50	1:00	1:50
WA	RV	1:00	1:30	2:30
ZP	MN	2:20	2:00	4:20
∑				16:50
Part 4	**cell assignment**	**IT**	**WT**	∑
TM	RV	1:00	1:00	2:00
BD	SF	0:20	0:50	1:10
SC	RN	1:30	2:45	4:15
TG	SV	1:20	0:40	2:00
RZ	SN	0:45	0.50	1:35
SL	RV	2:30	1:40	4:10
∑				15:10
Part 5	**cell assignment**	**IT**	**WT**	∑
AW	RF	1:10	2:15	3:25
PS	MV	1:20	1:15	2:25
BR	RN	1:00	3:30	4:30
BG	RF	1:45	2:10	3:55
FM	MF	1:35	1:30	2:05
∑				17:30
∑∑				81:15

Note. IT Instruction time
WT Working time

Appendix D: Academic Intelligence Standard Scores Based on the Normative Sample of the BIS Study 1 and 2 (Jäger et al., 1997) and Correlations with Age Study 2

(a) Standard scores of academic intelligence based on the normative sample of the BIS

	Sample Study 1 (N = 126)				Sample Study 2 (N = 182)			
	M	**SD**	**Min**	**Max**	**M**	**SD**	**Min**	**Max**
Cell SF	99.859	10.089	78.00	125.00	93.890	7.472	75.33	117.67
Cell SV	100.895	10.727	72.00	126.00	103.861	8.003	84.67	126.67
Cell SN	97.160	9.133	78.00	120.67	94.013	8.504	76.67	120.00
Cell MF	101.249	8.364	80.33	120.33	97.623	8.236	76.33	116.67
Cell MV	99.944	8.308	79.50	121.50	96.220	7.729	78.67	114.33
Cell MN	102.488	7.973	80.00	123.50	99.729	7.241	83.00	117.33
Cell RF	100.562	6.746	85.75	116.75	98.260	9.334	81.00	123.00
Cell RV	103.624	7.260	86.00	119.80	101.969	8.355	85.33	124.33
Cell RN	98.634	7.546	83.40	117.40	95.498	8.894	78.67	117.00
BIS-Speed	99.305	6.911	83.22	119.61	97.258	6.208	81.33	119.33
BIS-Memory	101.227	6.099	84.28	116.83	97.857	6.364	80.11	112.33
BIS-Reasoning	100.940	5.864	87.00	114.42	98.576	7.287	83.44	118.22
BIS-Verbal	101.488	6.558	85.23	116.10	100.683	6.057	87.89	118.89
BIS-Figural	100.557	5.554	88.31	114.39	96.594	6.062	84.11	112.44
BIS-Numerical	99.427	6.548	86.82	118.50	96.413	6.546	83.22	118.11
BIS-g	100.490	4.817	89.95	113.09	97.897	5.266	87.44	115.96

(b) Correlations with age prior to and after age standardization (age groups: 23-25; 26-30; 31-35; 36-40); N = 182, *p < .05, **p<.01

	Correlations with age	
	Prior to standardization	**After standardization**
Cell SF	-.083	.188*
Cell SV	.124	.064
Cell SN	.017	.204**
Cell MF	-.153*	-.002
Cell MV	-.166*	.054
Cell MN	-.153*	.049
Cell RF	-.316**	.051
Cell RV	-.193**	.055
Cell RN	-.112	.164*
BIS-Speed	.028	.197**
BIS-Memory	-.193**	.041
BIS-Reasoning	-.254**	.111
BIS-Verbal	-.102	.075
BIS-Figural	-.265**	.109
BIS-Numerical	-.102	.178*
BIS-g	-.182*	.145

Appendix E: Example Picture Sequence and Item for SMp2 Study 2

Example 1: Sort the following persons according to their seating arrangement at the dinner table in the last picture of the sequence. Start at the left hand in the front (free response).

1


2


3


4


Example 2: Which of the following extracts was NOT part of the picture sequence?

1


2


3


4


5


Appendix F: Course of Testing in Study 2 Including Planned Task Durations

Study 2 – Testing day 1

No.	Test	Continue program with	Planned Duration / min.
1	General instruction		10
	Biographical Questionnaire		
2	SRT	Q / Q	6
3	Word span	Q / Q	10
4	SPf1	U / A	13
5	Recognition of Repeated Tones	R / K	12
6	Instruction to Scenarios		20
	Example Scenario: Birger	S	
7	Scenario 1: Renate (SU_RF)	P / E	25
8	Questionnaire: Exhaustion		1
9	Break		10
10	BIS Part 1		17
11	SPa1	I / S	15
12	SMp1	E / S	12
13	SPf2	C / H	13
14	Memory Updating – Numerical	Q / Q	10
15	Scenario 2: Bringfried (SU_BS)	M / E	20
16	Break		5
17	BIS Part 2		15
18	Masked Words	C / K	8
19	SMv1		14
20	Dissected sentences	T / G	12
21	Scenario 3: Conny (SU_CK)	A / N	20
22	Questionnaire: Empathy		5
23	Questionnaire: Depression		5
24	Break		10
25	Readspeed (RS)	Z / W	10
26	SPv2	U / N / D / E	8
27	SMp2	R / B	20
28	Scenario 4: Christoph (SU_CP)	A / R	20
29	Questionnaire: NEO-FFI		10
30	Questionnaire: Exhaustion		1
	Sum		367

Study 2 – Testing day 2

No.	Test	Continue program with	Planned Duration / min.
1	Mouse Task (MT)	Q / Q	6
2	SPp1	U / E	8
3	SMa2	L / L	20
4	PONS Video Version (PONS-V)	W	12
5	Dot span	Q / Q	10
6	Scenario 5: Katharina (SU_KL)	A / S	20
7	Questionnaire: Exhaustion		1
8	Break		5
9	BIS Part 3		25
10	Rhythm	S	6
11	SMv2		13
12	SPa2	E	13
13	Tonal series	R / I	10
14	Scenario 6: Friedrich (SU_FB)	S / T	20
15	Break		20
16	BIS Part 4		22
17	SMa1	G	15
18	SPp2	E / S	12
19	Audiobook	U	9
20	Scenario 7: Hannah (SU_HR)	N / D	20
21	Questionnaire: AES		7
22	Break		10
23	SPv1	U / N / D / S	14
24	Scenario 8: Matthias (SU_MM)	O	20
25	SMf1+2	K / A	22
26	Questionnaire: Social Behavior / Altruism		10
27	Long Term Memory (LTM)	L / T	15
28	Questionnaire: Exhaustion		1
	Sum		367

BIS-Tasks (selection and order, 4 test parts)

Part 1	**cell assignment**	**IT**	**WT**	**∑**
VS	--	0:40	1:30	2:10
SI	SN	0:20	0:50	1:10
WM	MV		0:40 1:30	2:10
RD	RN	0:40	3:20	4:00
ZS	SF	0:30	1:00	1:30
ZZ	MN		1:00 0:50	1:50
WS	RV	0:30	1:00	1:30
∑				14:20

Part 2	**cell assignment**	**IT**	**WT**	**∑**
WE	MF		0:30 0:40	1:10
UW	SV	0:20	0:50	1:10
AN	RF	1:00	1:45	2:45
XG	SN	0:50	1:00	1:50
WA	RV	1:00	1:30	2:30
ZP	MN	0:20	2:00 2:00	4:20
BD	SF	0:20	0:50	1:10
∑				14:55

Part 3	**cell assignment**	**IT**	**WT**	**∑**
ZN	RN	0:50	3:40	4:30
ST	MV		1:00 2:00	3:00
RZ	SN	0:45	0:50	1:35
TM	RV	1:00	1:00	2:00
XG_new	SN – new task	0:50	1:00	1:50
OG	MF		1:30 1:40	3:10
TG	SV	1:20	0:40	2:00
CH	RF	1:00	3:00	4:00
∑				22:05

Part 4	**cell assignment**	**IT**	**WT**	**∑**
FM	MF	0:45	0:50 1:30	3:05
SC	RN	1:30	2:45	4:15
OE	SF	0:30	0:30	1:00
ZW	MN		1:00 2:00	3:00
AW	RF	1:10	2:15	3:25
SI_new	SN – new tasks	0:20	0:50	1:10
KW	SV	0:15	0:30	0:45
PS	MV	0:20	1:00 1:15	2:35
∑				19:15
∑∑				68:35

Note. IT Instruction time
WT Working time

Appendix G: Correlations of Social Perception Tasks and Speed Baseline Measures With BIS Cells and Within the Baseline Measures and the BIS Cells

(a) Correlations of standardized residuals of social perception tasks with social understanding

	SPv1_res	SPv2_res	SPa1_res [a]	SPa2_res	SPp1_res	SPp2_res	SPf1_res[b]	SPf2_res
SUv	.020	-.039	.037	.102	.102	-.014	.154*	.096
SUa	-.003	-.027	-.112	-.013	-.016	-.040	-.060	.011
SUp	-.032	-.095	-.071	-.030	-.071	-.116	-.011	-.055
SUf	-.020	-.067	.004	.071	-.011	-.026	.141	.072

Note. [a] N = 181; [b] N = 177; _res = standardized residual after controlling for the baseline variance

(b) Intercorrelations within the BIS cells and of the BIS cells with the speed baseline measures and the original social perception tasks

	BIS cells								
	RV	**RF**	**RN**	**MV**	**MF**	**MN**	**SV**	**SF**	**SN**
RF	.470**								
RN	.420**	.622**							
MV	.330**	.191**	.332**						
MF	.214**	.374**	.337**	.462**					
MN	.190*	.205**	.338**	.566**	.482**				
SV	.364**	.150*	.247**	.353**	.261**	.257**			
SF	.199**	.145	.092	.152*	.282**	.180*	.433**		
SN	.459**	.467**	.686**	.265**	.236**	.267**	.496**	.257**	
SRT	-.197**	-.158*	-.111	-.139	-.075	-.095	-.278**	-.343**	-.198**
MT	-.189*	-.323**	-.211**	-.117	-.132	-.048	-.087	-.126	-.115
RS	-.488**	-.226**	-.313**	-.242**	-.224**	-.182*	-.509**	-.193**	-.442**
SPv1	-.430**	-.206**	-.194**	-.261**	-.228**	-.210**	-.394**	-.175*	-.261**
SPv2	-.475**	-.228**	-.190*	-.300**	-.190*	-.215**	-.465**	-.229**	-.310**
SPa1	-.089	-.093	.047	-.020	-.107	-.047	-.167*	-.203**	-.030
SPa2[a]	-.227**	-.187*	-.079	-.135	-.157*	-.008	-.204**	-.220**	-.043
SPp1	-.121	-.183*	-.037	-.120	-.237**	-.234**	-.170*	-.393**	-.081
SPp2	-.060	-.197**	-.131	-.022	-.152*	-.027	-.050	-.176*	-.055
SPf1[b]	.022	-.009	.041	-.025	-.237**	-.085	-.007	-.222**	.084
SPf2	.051	-.032	-.065	-.119	-.096	-.070	-.108	-.077	-.007

(c) Intercorrelations within speed baseline tasks and with the original social perception tasks

	SRT	MT	SPv1	SPv2	SPa1 [a]	SPa2	SPp1	SPp2	SPf1[b]	SPf2
SRT			.303**	.322**	.433**	.344**	.421**	.241**	.198**	.248**
MT	.313**		.148*	.221**	.194**	.167*	.491**	.196**	.088	.112
RS	.309**	.181*	.627**	.792**	.218**	.297**	.173*	.175*	.201**	.139

Note. [a] N = 181; [b] N = 177

Appendix H: Correlations Within the Personality Trait Inventories in Study 2

	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C	Altruism	EC	PT
NEO-E	-.549**							
NEO-O	.092	-.039						
NEO-A	-.207**	.324**	-.108					
NEO-C	-.425**	.262**	-.046	.191**				
Altruism	.045	.212**	.248**	.264**	.041			
EC	.055	.095	.124	.336**	.003	.489**		
PT	-.086	.115	.229**	.321**	.159*	.207**	.130	
Depression	.815**	-.613**	.065	-.263**	-.532**	-.050	.057	-.183*

N = 182, * p < .05, ** p < .01

Appendix I: Kurzfassung zur Dissertation in deutscher Sprache

Kurzfassung zur Dissertation in deutscher Sprache

zum Thema „*Theorie und Messung Sozialer Intelligenz als Kognitives Leistungskonstrukt*“, vorgelegt von Dipl.-Psychologin Susanne Weis

Die Erforschung menschlicher Fähigkeiten blickt auf eine lange und erfolgreiche Forschungstradition zurück. Dabei steht das Fähigkeitskonstrukt der akademischen Intelligenz unangefochten im Zentrum der Forschung. Gleichzeitig fungiert es als Maßstab, an dem sich alle weiteren Kandidaten für ein neues Fähigkeitskonstrukt messen sollten. Das übergeordnete Ziel der vorliegenden Dissertation ist es, die Forschung zu sozialer Intelligenz als Kandidat für ein neues Fähigkeitskonstrukt theoretisch und methodisch aufzuarbeiten und weiterzuentwickeln. Dies geschieht unter Bezugnahme auf in der Literatur genannte konsensuale Anforderungen an ein neues Fähigkeitskonstrukt (Matthews, Zeidner & Roberts, 2005; O'Sullivan, 1983; Schaie, 2001; Süß, 2001, 2006; Weber & Westmeyer, 2001). Dazu gehören eine a-priori Begriffsklärung, theoretische Einordnung in das nomologische Netzwerk bereits etablierter Konstrukte und die Beschränkung auf hinreichend generelle und kognitive Fähigkeiten. Weitere Kriterien betreffen die methodischen Voraussetzungen. Dabei steht die Operationalisierung anhand von objektiven Leistungsdaten sensu Cattell (1965) im Mittelpunkt. Außerdem soll der Nachweis konvergenter und divergenter Konstruktvalidität erbracht werden. Nicht zuletzt wird der Nachweis der inkrementellen Bedeutsamkeit bei der Vorhersage relevanter Außenkriterien verlangt (über etablierte Konstrukte hinaus). Vor diesem Hintergrund kann die soziale Intelligenz bislang nicht als etabliertes Fähigkeitskonstrukt angesehen werden

Die spezifischeren Ziele dieser Arbeit lassen sich grob in drei Bereiche untergliedern. (1) Die theoretischen und methodischen Grundlagen sozialer Intelligenz als kognitives Fähigkeitskonstrukt sollen erarbeitet und dargelegt werden. (2) Diesen Überlegungen folgend soll eine Testbatterie auf Basis von Leistungstests entwickelt werden. (3) In zwei ersten empirischen Untersuchungen sollen die psychometrischen Eigenschaften der Skalen untersucht und erste Validierungsbefunde berichtet werden.

Theoretische und methodische Grundlagen

Die vorliegende Arbeit bezieht sich auf das Leistungsmodell sozialer Intelligenz von Weis und Süß (2005) (siehe auch Weis, Seidel & Süß, 2006) und auf ein integratives Modell sozial kompetenten Verhaltens von Süß, Weis und Seidel (2005). Soziale Intelligenz wird als multidimensionales Fähigkeitskonstrukt definiert und klar von verhaltensbasierten Konzeptionen abgegrenzt. Jenes Leistungsmodell setzt sich zusammen aus den kognitiven Operationen soziales Verständnis, soziales Gedächtnis, soziale Wahrnehmung und soziale Kreativität. In der Literatur wird häufig soziales Wissen als weitere kognitive Fähigkeitskomponente angeführt (Cantor & Harlowe, 1994; Lee, Day, Meara & Maxwell, 2002). Allerdings ist die Rolle von Wissen in Intelligenzkonstrukten umstritten, da es kultur- und kontextabhängig ist (Süß, 1996; Weis et al., 2006) und somit nicht allein auf kognitiven Anforderungen basiert. Aus diesem Grund wurde bei der nachfolgenden Testkonstruktion soziales Wissen nicht berücksichtigt, ebenso wie soziale Kreativität - vor allem aus Aufwandsgründen - nicht in die aktuelle Testkonstruktion aufgenommen wurde.

In Erweiterung zum Kernmodell von Weis und Süß (2005) erarbeitet die vorliegende Dissertation weitere taxonomische Einteilungsgesichtspunkte, die aus Definitionen sozialer Intelligenz und verwandter Konstrukte abgeleitet sind. Der Wert einer solchen Taxonomie ist dreigeteilt: Zum ersten wird damit die Definition des Konstrukts auf eine heterogenere und repräsentativere theoretische Basis gestellt (Cattell, 1987), zum zweiten werden bislang unberücksichtigte Varianzquellen bei der Operationalisierung des Konstrukts systematisch kontrolliert, zum dritten ergeben sich aus solchen taxonomischen Überlegungen testbare Hypothesen über die Struktur eines Konstrukts, die möglicherweise zur Etablierung eines facettentheoretischen Modells beitragen. Mit den vorliegenden taxonomischen Überlegungen werden folgende Elemente unterschieden: (a) materialgebundene Inhalte oder damit verbundene soziale Stimuli (z.B. Körpersprache dargestellt in Videos, der Tonfall der Stimme in Tonaufzeichnungen, etc.), (b) die abgefragte Modalität (z.B. Emotionen oder Gedanken einer Person, über die in einer sozialen Situationen geschlussfolgert werden sollen), (c) das Setting (z.B. der Kontext, wie etwa beruflich oder privat) und (d) die Target- oder Zielpersonen, deren Emotionen oder Gedanken geschlussfolgert werden sollen. Die Taxonomie erhebt dabei keinen Anspruch auf Vollständigkeit und muss empirisch überprüft werden.

Bei der in dieser Arbeit vorgestellten Aufarbeitung methodischer Grundlagen wird der Fokus allein auf Leistungstests gelegt, die der objektiven Erfassung sozialer Intelligenz und verwandter Konstrukte dienen. Vorherrschende methodische Fragestellungen und Probleme werden dargestellt und bestehend Mess- und Testansätze vor diesem Hintergrund diskutiert. Dabei werden sowohl formale als auch inhaltliche Probleme identifiziert wie (a) die Auswahl adäquater Item- und Antwortformate, (b) die Verwendung artifizieller, dekontextualisierter und realitätsferner Stimuli und (c) der fehlende theoretische Bezug bei der Testkonstruktion und die daraus folgende mangelnde Übereinstimmung zwischen Theorie und Messkonstrukt.

Entwicklung des Sozialen Intelligenztests Magdeburg (SIM)

Basierend auf dem Leistungsmodell sozialer Intelligenz nach Weis und Süß (2005) und den bereits genannten weiterführenden taxonomischen Überlegungen wurde der Soziale Intelligenztest Magdeburg (SIM) entwickelt. Das Testdesign beinhaltete drei operative Fähigkeitsbereiche (soziales Verständnis, Gedächtnis und Wahrnehmung) und vier material-gebundene Inhaltsbereiche (geschriebene und gesprochene Sprache, Bilder und Videos). Zusätzlich wurden bei allen Aufgaben sowohl das Setting der dargestellten Situationen als auch die Anzahl der dargestellten Personen systematisch variiert. Alle Aufgaben basieren auf realem Aufgabenmaterial. Im Folgenden soll der jeweilige Testansatz der einzelnen operativen Fähigkeitsbereiche kurz dargestellt werden:

a) soziales Verständnis

Soziales Verständnis wurde basierend auf einem Szenarioansatz operationalisiert. Jedes Szenario setzte eine Zielperson ins Zentrum der Aufgaben, deren Emotionen, Gedanken, Beziehungen zu Dritten und Persönlichkeitseigenschaften vom Probanden eingeschätzt werden sollten. Die Einschätzung erfolgte auf Basis von Aufgabenmaterialien, die den Inhaltsbereichen zugeordnet waren (z.B. Briefe oder Emails, Ton- oder Videoaufzeichnungen von Gesprächen und anderen sozialen Interaktionen). Antwortformat war eine 7-stufige Ratingskala, auf der die Probanden einschätzten, wie stark beispielsweise eine Emotion bei der Zielperson ausgeprägt war (von 1 = gar nicht bis 7 = sehr stark). Die Antworten der Probanden wurden anhand des sogenannten Targetscoring ausgewertet (der gewichtete Betrag der Differenz zwischen der Antwort der Probanden und der Antwort der Zielperson).

b) soziales Gedächtnis

Die Aufgaben zum sozialen Gedächtnis beinhalteten die zeitlich limitierte Darbietung von sozialen Stimuli mit Hilfe der bereits genannten Aufgabenmaterialien. Probanden waren aufgefordert sich sozial relevante Inhalte einzuprägen und später im Multiple-Choice-Format oder in ungebundenen Antwortformaten wiederzugeben. Die Antwortzeit war ebenfalls limitiert. Die Antworten wurden entsprechend des Prozentsatzes an richtigen Antworten ausgewertet.

c) soziale Wahrnehmung

Aufgaben zur sozialen Wahrnehmung bedienten sich ebenfalls des „Targetkonzepts“. Anders als in den Aufgaben zum sozialen Verständnis waren Targets hier allgemein als Zielreize anzusehen (z.B. eine vorgegebene geschriebene oder gesprochene Aussage wie ein ausgesprochener Dank oder eine Zustimmung, eine bestimmte Körperbewegung oder eine Interaktion beispielweise in Form von Augenkontakt). Aufgabe der Probanden war es, vorgegebene Zielreize im Stimulusmaterial so schnell wie möglich zu erkennen und diese Wahrnehmung durch einen Tastendruck anzuzeigen. Das abhängige Maß war die Reaktionszeit richtiger Antworten.

Die finale Testbatterie beinhaltete acht Szenarien, sechs Aufgaben im Bereich des sozialen Gedächtnisses und zwei Aufgaben pro Zelle aus dem Bereich der sozialen Wahrnehmung.

Methoden und Studien

Die der Arbeit zugrunde liegenden Fragestellungen betrafen die psychometrischen Eigenschaften der neu entwickelten Skalen und die Konstruktvalidität der sozialen Intelligenz - operationalisiert durch den SIM. Hypothesen zur Konstruktvalidierung wurden in Form von Strukturmodellen aufgestellt. Strukturmodelle erlauben die Überprüfung der Passung eines theoretischen Modells zu der empirischen Datenstruktur mit Hilfe von konfirmatorischen Faktorenanalysen.

Die vorliegende Dissertation umfasste zwei empirische Studien. An Studie 1 nahmen 126 Studenten der Otto-von-Guericke-Universität Magdeburg teil, im Durchschnitt 21.35 Jahre alt (*sd* = 3.06), wobei 53.5 % der Stichprobe weiblich waren. Die Stichprobe in Studie 2 setzte sich aus 182 Erwachsenen zusammen, heterogen im Hinblick auf Alter, Bildung und Beruf. Die Probanden waren durchschnittlich 28.69 Jahre alt (*sd* = 5.57), 58.8 % waren weiblich. In beiden Studien wurde der SIM in seiner jeweils aktuellen Version eingesetzt, zudem der Berliner Intelligenz Strukturtest (BIS-Test; Jäger, Süß & Beauducel, 1997) und ein Fragebogen zur Erfassung der Big Five Persönlichkeitsfaktoren. Überdies wurde eine Vielzahl weiterer Instrumente eingesetzt, die nicht direkt mit den zentralen Fragestellungen in Zusammenhang stehen und hier nicht weiter beschrieben werden sollen.

Ergebnisse

Der erste Teil der Analysen befasste sich mit den psychometrischen Eigenschaften der neu entwickelten Skalen. Diese zeigten weitestgehend ausreichend bis gute psychometrische Qualität. Die Reliabilität der Skalen (Cronbach’s Alpha) zum sozialen Verständnis lag zwischen .75 und .85, zum sozialen Gedächtnis bei .19 bis .84 (mit niedrigen Werten für eine Aufgabe auf Basis von gesprochener Sprache (.19) und zwei Aufgaben auf Basis von Bildern (.46 und .56)). Die sozialen Wahrnehmungsaufgaben zeigten mit .71 bis .98 die höchsten Reliabilitätswerte. Alle Skalen erwiesen sich als annähernd normalverteilt.

Zur Frage der internen Struktur der sozialen Intelligenz: Konfirmatorische Faktorenanalysen in beiden Studien bestätigten die multidimensionale Struktur sozialer Intelligenz mit zwei korrelierten operativen Faktoren (soziales Verständnis und soziales Gedächtnis). Die Korrelation der Faktoren lag bei r = .35 in Studie 1 und r = .20 in Studie 2. Im Gegensatz zur ersten Studie konnte bei der zweiten Untersuchung kein übergeordneter Faktor soziale Intelligenz identifiziert werden. Damit stellte sich für weitere Forschung die Frage, ob nach Abschluss noch ausstehender Schritte der Testentwicklung soziale Intelligenz als einheitliches Konstrukt weiter bestehen kann.

Zur Frage der divergenten Konstruktvalidierung: Korrelative Analysen und konfirmatorische Faktorenanalysen bestätigten die Unabhängigkeit der sozialen Fähigkeitsfaktoren von denen der akademischen Intelligenz. Zwar zeigte vor allem der Faktor soziales Gedächtnis substantielle Korrelationen mit dem Gedächtnisfaktor des BIS-Tests (r = .42 / .67 in Studie 1 / 2), allerdings konnten weiterführende Analysen zeigen, dass sowohl die Struktur der sozialen Intelligenz als auch die Korrelationen ihrer operativen Fähigkeitsbereiche erhalten blieben, wenn BIS-Varianz auspartialisiert wurde. Die sozialen Intelligenzaufgaben zeigten sich außerdem als weitgehend unkorreliert mit den Big Five Persönlichkeitsfaktoren.

Diskussion

Neben der Diskussion der eben angerissenen empirischen Ergebnisse zu psychometrischen Eigenschaften der Skalen und der Konstruktvalidierung bezieht sich die Diskussion vor allem auf den gewählten Testansatz, der mehrere diskutable Entscheidungen beinhaltet. Diese betreffen zunächst den Gebrauch von realistischem Aufgabenmaterial, danach die verwendete Scoringmethode für die Aufgaben des sozialen Verständnisses, den eigentlichen Konstruktionsprozess und abschließend der Lösung der sozialen Verständnisaufgaben zugrunde liegende Prozesse. Zukünftige Fragestellungen und Untersuchungsdesigns, die sich aus der vorliegenden Arbeit ergeben, werden ebenfalls aufgezeigt.

Appendix J: Wissenschaftlicher Werdegang

Angaben zur Person

Geb. am:	27.08.1974
Geb. in:	Frankenthal / Pfalz
Familienstand:	ledig
Staatsangehörigkeit:	deutsch
Privatadresse:	Anton-Bruckner-Str. 6c 67240 Bobenheim-Roxheim
Telefon privat:	06239 / 926410
Email	susanne.weis@ovgu.de

Schulische Ausbildung

1981-1985	Friedrich-Ebert-Grundschule Frankenthal
1985-1994	Albert-Einstein-Gymnasium Frankenthal
18.06.1994	Abitur am Albert-Einstein-Gymnasium Frankenthal/Pfalz mit der Note 1.3

Akademische Ausbildung

1994-1995	Philipps-Universität Marburg: 2 Semester Jurastudium
Ab Okt. 1995	Psychologiestudium an der Universität Mannheim
06.11.1997	Vordiplom
Dez. 1998 bis Juni 2001	Stipendium bei der Studienstiftung des Deutschen Volkes
04.04.2002	Diplomarbeit zum Thema „Facets of Social Intelligence – Cognitive Performance Measures in a Multitrait-Multimethod Design" mit der Note ‚sehr gut'
18.10.2002	Abschluss: Diplom mit der Note ‚sehr gut'
März 2007	Abschluss des curriculums "Sportpsychologie im Leistungssport" mit dem Abschluss "Sportpsychologin asp/bdp"
August 2007	geplante Abgabe der Dissertationsschrift zum Thema ‚Social Intelligence - Theoretical Substantiation and Methodological Re-Invention of an Aged and Misconceived Performance Construct'

Berufliche Tätigkeiten

Zw. Juli 1996 und August 1997	Dreimonatiges Forschungspraktikum am Forschungsprojekt "Arbeitsgedächtnis und Intelligenz", Universität Mannheim
Nov. 1996 bis Juni 1999 und Apr.-Dez. 2001	Wissenschaftliche Hilfskraft am Forschungsprojekt "Arbeitsgedächtnis und Intelligenz" und am Lehrstuhl Psychologie II (Methodenlehre) an der Universität Mannheim
Okt. 1997 bis Juli 1998	Durchführung des vorlesungsbegleitenden Statistik-Tutoriums
Nov. 1999 bis Jan. 2000	Praktikum beim Beratungsinstitut Liebelt in Würzburg
Nov.-Dez. 2002 Feb.-März 2003 Juni-Aug. 2003 Okt.-Nov. 2003	Geprüfte wissenschaftliche Hilfskraft am Lehrstuhl Methodenlehre, Psychodiagnostik und Evaluationsforschung an der Otto-von-Guericke-Universität Magdeburg
Nov. 2003 bis März 2005	Wissenschaftliche Mitarbeiterin am Forschungsprojekt "Kognitive Facetten Sozialer Intelligenz" an der Otto-von-Guericke-Universität (OvGU) Magdeburg
April 2004	Ernennung zur Lehrbeauftragten durch OvGU Magdeburg
April 2005 bis März 2006	Vertretung der wissenschaftlichen Assistentenstelle (C1) in der Abteilung Methodenlehre, Psychodiagnostik und Evaluationsforschung der OvGU Magdeburg (halbe Stelle)
Seit April 2006	Vertretung der wissenschaftlichen Assistentenstelle (C1) in der Abteilung Methodenlehre, Psychodiagnostik und Evaluationsforschung der OvGU Magdeburg (volle Stelle)
Seit April 2007	Wissenschaftliche Mitarbeiterin in der Abteilung Methodenlehre, Psychodiagnostik und Evaluationsforschung der OvGU Magdeburg (volle Stelle; TV-L E13)