EMOTIONS AND PERFORMANCE IN VIRTUAL WORLDS

An Empirical Study in the Presence of Missing Data

Inauguraldissertation
zur Erlangung des Doktorgrades der
Wirtschafts- und Sozialwissenschaftlichen Fakultät der
Universität zu Köln

2015

vorgelegt von

Diplom-Informatikerin
Sabrina Schiele

aus Duisburg

Referent: Prof. Dr. Detlef Schoder
Koreferent: Prof. Dr. Claudia Loebbecke
Datum der Promotion: 27.11.2015

EMOTIONS AND PERFORMANCE IN VIRTUAL WORLDS

SABRINA SCHIELE

Department of Information Systems and Information Management
Faculty of Management, Economics and Social Sciences
University of Cologne

Sabrina Schiele: *Emotions and Performance in Virtual Worlds,* An Empirical Study in the Presence of Missing Data, 2015

To Gregor,
who supported all of my decisions and
always found the right words to keep me going
in times of despair.

ABSTRACT

In this work, we first investigate characteristics of virtual worlds and determine important situational variables concerning virtual world usage. Moreover, we develop a model which relates individual differences of virtual world users, namely emotional and cognitive abilities, experiences with virtual worlds as a child, and the level of cognitive absorption perceived during virtual world use, to the users' individual performance in virtual worlds. We further test our model with observed data from 4,048 study participants. Our results suggest that cognitive ability, childhood media experience, and cognitive absorption influence multiple facets of emotional capabilities, which in turn have a varyingly strong effect on virtual world performance among different groups. Notably, in the present study, the effect of emotional capabilities on performance was stronger for users which prefer virtual worlds that have more emotional content and require more social and strategic skills, particularly related to human behavior. Interestingly, while cognitive ability was positively related to various emotional capabilities, no evidence for a direct path between cognitive ability to performance could be identified. Similarly, cognitive absorption positively affected emotion perception, yet did not influence performance directly. Our findings make the case for abandoning the traditional perspective on IS–which mainly relies on mere usage measures–and call for a more comprehensive understanding and clearer conceptualizations of human performance in psychometric studies. Additionally, our study treats missing data (an inherent property of the data underlying our study), links their presence to theoretical and practical issues, and discusses implications.

ZUSAMMENFASSUNG

In der vorliegenden Arbeit untersuchen wir zunächst die charakteristischen Eigenschaften virtueller Welten und ergründen die besonderen Umstände ihrer Nutzung. Zudem entwickeln wir ein Modell, welches die Performanz von Nutzern virtueller Welten in Bezug setzt zu ihren emotionalen und kognitiven Fähigkeiten, ihrem Erfahrungshintergrund bezüglich virtueller Welten im Kindesalter, und dem Niveau an kognitiver Absorption, welches sie während der Nutzung von virtuellen Welten erleben. Desweiteren testen wir unser Modell anhand von Daten, welche wir im Rahmen unserer Studie mit 4.408 Teilnehmern erhoben haben. Unsere Ergebnisse deuten darauf hin, dass kognitive Fähigkeiten, Medienerfahrung in der Kindheit und kognitive Absorption die emotionalen Fähigkeiten von Nutzern virtueller Welten beeinflussen, und dass diese Wirkbeziehung wiederum die Performanz der Nutzer beeinflusst–jedoch je nach Nutzergruppe unterschiedlich stark. Insbesondere war die Wirkung von emotionalen Fähigkeiten auf Performanz in der vorliegenden Studie größer für diejenigen Nutzer, die virtuelle Welten bevorzugen, welche mehr emotionale Inhalte enthalten und deren Aufgabenstellung mehr soziale Kompetenzen und strategisches Geschick erfordern, vor allem bezüglich menschlicher Verhaltensweisen. Interessanterweise ließ sich kein direkter Zusammenhang zwischen Performanz und kognitiven Fähigkeiten nachweisen, auch wenn letztere einen wichtigen Einfluss auf verschiedene emotionale Fähigkeiten zeigten. Ähnlich verhielt es sich mit kognitiver Absorption, welche sich zwar auf die Wahrnehmung von Emotionen auswirkte, jedoch nicht direkt auf Performanz. Als Fazit unserer Untersuchung schlagen wir vor, die traditionelle Sicht auf Informationssysteme, welche hauptsächlich auf bloßen Nutzungsstatistiken fußt, aufzugeben, und im Hinblick auf zukünftige psychometrische Studien ein umfassenderes Verständnis und eine klarere Konzeptualisierung von menschlicher Performanz zu entwickeln. Zusätzlich befasst sich unsere Studie mit fehlenden Werten (welche unseren Daten in größerem Umfang anhafteten), untersucht damit verbundene theoretische und praktische Problemstellungen und diskutiert Implikationen.

LIST OF PUBLICATIONS

In reverse chronological order; publications C-E appeared under my maiden name Steinfels.

A. Weiss, T. & Schiele, S. (2013). Virtual worlds in competitive contexts: Analyzing eSports consumer needs. *Electronic Markets*, *23*(4), 307–316. doi:10.1007/s12525-013-0127-5.

B. Schiele, S., Weiss, T., & Putzke, J. (2011). On inter-reality literacy: Emotions as predictors of performance in virtual worlds [Research in progress]. In D. F. Galletta & T.-P. Liang (Eds.), *Proceedings of the Thirty-Second International Conference on Information Systems (ICIS 2011).*

C. Floeck, F., Putzke, J., Steinfels, S., Fischbach, K., & Schoder, D. (2011). Imitation and quality of tags in social bookmarking systems: Collective intelligence leading to folksonomies. In T. Bastiaens, U. Baumöl, & B. Krämer (Eds.), *On collective intelligence* (Vol. 76, pp. 75–91). Advances in Intelligent and Soft Computing. Berlin: Springer. doi:10.1007/978-3-642-14481-3_7.

D. Oster, D., Schoder, D., Putzke, J., Fischbach, K., Gloor, P. A., & Steinfels, S. (2010). Tell your customers what they really want to hear: Improving the effectiveness of advertising campaigns in the financial sector using SNA on the Web2.0. In *Proceedings of the 2010 International Sunbelt Social Network Conference (Sunbelt XXX).*

E. Oster, D., Steinfels, S., Putzke, J., Fischbach, K., Gloor, P. A., & Schoder, D. (2009). Measuring and enhancing advertising success using SNA on the Web. In *Proceedings of the 2009 International Sunbelt Social Network Conference (Sunbelt XXIX).*

Some ideas and figures presented in the following have been developed and published prior to this thesis as part of the publications above; references to them will be made in the relevant text passages accordingly.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

AI	artificial intelligence
ADF	asymptotic distribution free
AGFI	adjusted goodness-of-fit index
APA	American Psychological Association
AVE	average variance extracted
BSEM	Bayesian structural equation modeling
BTA	“better-than-average”
CBSEM	covariance-based structural equation modeling
CFA	confirmatory factor analysis
CFIT	Culture Fair Intelligence Test
CHC	Cattell-Horn-Carroll
CFI	comparative fit index
CIRME	childhood inter-reality media experience
df	degrees of freedom
DIC	deviance information criterion
EFA	exploratory factor analysis
EI	emotional intelligence
EM	expectation-maximization
FIML	full information maximum likelihood
GFI	goodness-of-fit index
HCI	human-computer interaction
IP	Internet protocol
IR	inter-reality
IRT	item response theory
IS	Information Systems
IT	information technology
KMO	Kaiser-Meyer-Olkin
M	sample mean
MAP	minimum average partial
MAR	missing at random
MCAR	missing completely at random
MCMC	Markov chain Monte Carlo
MDT	missing data technique
ML	maximum likelihood
MMOG	massively multiplayer online game
MNAR	missing not at random
MI	multiple imputation
MSA	measure of sampling adequacy
N	total number of cases
n	number of cases in a subsample
NFI	normed fit index
NNFI	nonnormed fit index

NFL	National Football League
PA	parallel analysis
PAF	principal axis factoring
PCA	principal component analysis
PLS	partial least squares
RMSEA	root mean square error of approximation
SD	standard deviation
SE	self-estimated
SEM	structural equation modeling
SRMR	standardized root mean square residual
SMC	squared multiple correlation
TAM	Technology Acceptance Model
TEIQUE	Trait Emotional Intelligence Questionnaire
TEIQUE-SF	Trait Emotional Intelligence Questionnaire-Short Form
TLI	Tucker-Lewis index
URL	uniform resource locator
USM	universal structure modeling
VR	virtual reality
WAIS-IV	Wechsler Adult Intelligence Scale-Fourth Edition
WPT	Wonderlic Personnel Test
XBA	cross-battery assessment approach

1 INTRODUCTION

This thesis covers two quite different subject areas, performance in virtual worlds on the one hand, missing data on the other. More specifically, the first subject area we will be dealing with revolves around topics in the field of Information Systems (IS), media, and virtual worlds and investigates explanatory variables for individual performance of virtual world users. Virtual worlds are computer-simulated three-dimensional environments that users experience with the aid of animated characters, so-called avatars (Animesh, Pinsonneault, Yang, & Oh, 2011). These environments are of academic and practical significance (Wasko, Teigland, Leidner, & Jarvenpaa, 2011) because their technology can "dramatically change how people interact, navigate Web sites, and conduct business" (Mennecke et al., 2008, p. 372). While initially designed for entertainment purposes (Bartle, 2004), virtual worlds have attracted a great deal of attention after showing potential for applications in business, educational, and government settings (cf. Schultze, 2010; Schultze & Orlikowski, 2010). Our review of relevant literature lead to the assumption that virtual worlds not only challenge users with regard to their technical or cognitive skills, but also with regard to their emotional capabilities. A research model accounting for emotions thus seemed to provide opportunities for explaining performance in a virtual world context. Theoretical foundations of the approach we eventually applied partly build on findings of the meta-analysis performed by Joseph and Newman (2010), who investigated the causal mechanisms of job performance, cognitive ability, and certain individual differences with respect to emotions. Our approach integrates some of the knowledge gained from these findings and transfers it into a self-contained study in a virtual world setting. We presented our ideas and preliminary considerations as a research-in-progress paper at the 32nd International Conference on Information Systems (ICIS 2011, held in Shanghai, China from December 4 to 7) at an early stage of our investigations:

> Schiele, S., Weiss, T., & Putzke, J. (2011). On inter-reality literacy: Emotions as predictors of performance in virtual worlds [Research in progress]. In D. F. Galletta & T.-P. Liang (Eds.), *Proceedings of the Thirty-Second International Conference on Information Systems (ICIS 2011)*.[1]

The actual study with approximately 6,600 respondents was carried out in the context of competitive online computer gaming (also referred to

1 Retrievable from http://aisel.aisnet.org/icis2011/proceedings/visualmedia/4.

as electronic sports or eSports), and data were[2] collected by means of a survey and a web crawler.

The second subject area of this thesis addresses missing data, a problem frequently encountered across all research disciplines (e. g., Kamakura & Wedel, 2000; Sinharay, Stern, & Russell, 2001; D. B. Rubin, 1996; Schafer & Olsen, 1998; Raykov, 2012; Roth, 1994). Dealing with missing data is important because their occurrence can seriously affect the validity of study findings (e. g., Allison, 2003; Schafer & Graham, 2002; McKnight, McKnight, Sidani, & Figueredo, 2007). We approach the subject by investigating the reasons for missing values, their possible impact on findings of a study, and their adequate treatment; in a less detailed manner, we also address related issues like data distributions and sample size. We did not initially plan to incorporate these topics. However, they eventually became an integral part of the present study after the very first examinations of our data foreshadowed the challenges that would stem from this domain.

Our study mainly targets the IS community, but also aims at informing practitioners and researchers in other disciplines (cf. G. Gill & Bhattacherjee, 2009). Due to the characteristics of the present study, the motivation of our research is twofold, and accordingly, the aim of our work can be broken down into two distinct objectives. In the following, we first elaborate on the motivation behind our interest in virtual worlds and state our related objective. Subsequently, we give more details on why missing values (together with other data issues) became such an important subject of this work and explain the aim we pursued in that field. We then proceed to an outline of the course of the study.

1.1 SUBJECT AREA I: VIRTUAL WORLDS

1.1.1 *Motivation*

The call for proposals for the Advanced Digital Gaming/Gamification Technologies research funding grant–part of the Horizon 2020 research and innovation framework program of the European Union–has delineated challenges and prospective gains related to digital games as follows:[3]

> Digital games and gamification mechanics applied in non-leisure contexts is an important but scattered industry that can bring high pay-offs and lead to the emergence of a prospering market. Digital games can also make a real change in the life of a large number of targeted excluded groups,

2 For a discussion concerning the question "Data are or data is?", see related entry of the The Guardian Blog, http://www.theguardian.com/news/datablog/2010/jul/16/data-plural-singular.

3 Official web page: http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/90-ict-21-2014.html, publication date: 2013-12-11.

> enhancing their better integration in society. This requires however the development of new methodologies and tools to produce, apply and use digital games and gamification techniques in non-leisure contexts, as well as building scientific evidence on their benefits–for governments, enterprises and individuals.

A very special form of digital games has moved to the center of interest: virtual worlds. Second Life, World of Warcraft, Minecraft, and other representatives of this genre are being perceived as promising and cost-effective means for communication, multimedia meetings, distributed collaboration, collaborative design, learning, training, networking, business development, and shopping as well as for real-time simulations in high-energy physics, surgery, and so forth (Goel, Johnson, Junglas, & Ives, 2011; Barnett & Coulson, 2010; Piccoli, Ahmad, & Ives, 2001; Riedl, Mohr, Kenning, Davis, & Heekeren, 2011; Blümel, Termath, & Haase, 2009; Powell, Piccoli, & Ives, 2004; Mennecke et al., 2008; Rosser et al., 2007; Petrakou, 2010; Dannecker et al., 2008; Annetta, 2010; Suh, Kim, & Suh, 2011; Gintautas & Hübler, 2007; C. Wagner, Schill, & Männer, 2002). Some even anticipate opportunities for the care of elderly people (Gee, Browne, & Kawamura, 2005), health improvement (Kato, 2010), rehabilitation (Nguyen, Merienne, & Martinez, 2010; cf. also Suh et al., 2011), psychotherapy (Barnett & Coulson, 2010), and the education of children with special needs (Kientz, Hayes, Westeyn, Starner, & Abowd, 2007).

Research on virtual worlds has its roots in engineering (Fetscherin, Lattemann, & Lang, 2008), thus has traditionally focused on technological issues (Animesh et al., 2011). Later, psychology particularly concentrated on the negative effects of media consumption (Przybylski, Rigby, & Ryan, 2010; Schiele et al., 2011). As a result, the additional value that the use of virtual worlds brings for different application fields as well as related risks is not yet sufficiently understood (Berente, Hansen, Pike, & Bateman, 2011; Clayes & Anderson, 2007; A. Davis, Murphy, Owens, Khazanchi, & Zigurs, 2009). It has been suggested that for virtual worlds to become more than just a demonstration of technical feasibility, a corporate strategy needs to be well-defined with respect to their purpose and application (cf. Mennecke et al., 2008; Jarvenpaa, Leidner, Teigland, & Wasko, 2007; Gartner, Inc., 2007). Messinger et al. (2009) proposed to classify virtual worlds by accounting for five key dimensions, namely for a virtual world's (a) purpose and content, (b) place and location, (c) platform and design of interaction, (d) population and interaction patterns as well as for its (e) profit model (cf. also C. E. Porter, 2004). Many companies already utilize virtual worlds as a "platform to reach consumers" (Wasko et al., 2011, p. 645); they have experimented with reverse product placement, that is, the transfer of fictional brands or products from the virtual into the real world (Edery, 2006), and vice versa, have created new demands for products that only exist in the virtual, so-called virtual goods (Animesh et al., 2011; Castronova, 2005).

The impression yet remains that despite all optimism expressed (cf. Wasko et al., 2011) and encouraging signs showing from early adopters (Nevo, Nevo, & Carmel, 2011), virtual worlds are far from being a resounding success outside the leisure context (Yang, Lim, Oh, Animesh, & Pinsonneault, 2012). The fundamental issue here is whether this is due to a limited understanding of the phenomenon and the consequent inability to unlock the full potential of virtual worlds, or alternatively, whether this reflects the fact that expectations concerning the usefulness of this technology in nonleisure contexts are simply too high.

Understanding the use of an information technology (IT) artifact is a key success measure for development and implementation processes (Barki, Titah, & Boffo, 2007). However, by exploring mechanisms that have the potential to explain user performance in virtual worlds, we take a different perspective to advance understanding of virtual worlds. Aspects like intentions to use virtual worlds (Goel et al., 2011), their acceptance as a technology (Holsapple & Wu, 2007; Fetscherin & Lattemann, 2008), and factors that foster (Choi & Kim, 2004) or hinder (Berente et al., 2011) their (continuous) use are thus not the focus of our interest. They do, however, play a supporting role in our effort to answer the following question: Given that people engage in virtual worlds–for reasons that are relatively well understood nowadays (cf. e. g., Nah, Eschenbrenner, & DeWester, 2011; Hsu & Lu, 2004; Wu, Li, & Rao, 2008; D. Williams, Yee, & Caplan, 2008)–what individual differences are capable of predicting these people's virtual world performance? Through this approach, we aim at substantially extending theoretical knowledge of processes inherent to virtual worlds, but also at making a practically relevant contribution, the latter of which may eventually conduce to the success of virtual worlds especially in serious application fields (cf. W. W. Chin & Marcolin, 2001; Barki et al., 2007).

Numerous appeals to study virtual worlds extensively for the purpose of examining social, behavioral, and economic issues–which appear to be "as complicated as those in the real world" (Animesh et al., 2011, p. 806)–have been made in the past. According to these appeals, virtual worlds are not only the object of investigation, but also serve as "virtual laboratories to explore aspects of human behavior" (Barnett & Coulson, 2010, p. 167). They invite researchers to compare "intra-world and inter-world practice" (Bray & Konsynski, 2007, p. 17), to observe interactions between artificial intelligence (AI) and humans or "even modeling them on specific human individuals to better understand the cognitive processes that shape human behavior" (Bainbridge, 2007, p. 475) in order to examine how virtual worlds are "used and misused by users, and [ultimately] the impact that they have on users, communities, organizations, and societies at large" (Mennecke et al., 2008, p. 371).

Although all kinds of disciplines such as IS, neurobiology, political science, organizational governance research, sociology, social psychology, and population ecology are called upon to participate, IS researchers seem to be in a particularly good position to analyze virtual worlds (cf. Agarwal & Lucas, 2005): Not only are they familiar with technical issues like design principles (Chaturvedi, Dolk, & Drnevich, 2011) or AIs–the latter of which are still rather simple than lifelike to this day (Bainbridge, 2007; Riedl et al., 2011; MacDorman, 2006a)–but they are also able to deal with social and business aspects of virtual worlds (cf. Animesh et al., 2011). As literature particularly emphasizes the importance of examining the outcome of virtual world use (Suh et al., 2011; Schultze & Orlikowski, 2010; Annetta, 2010), we aim to make a contribution to this stream of research, as explained in more detail below.

1.1.2 *Research Objective*

Research has begun to examine various kinds of outcomes of virtual world use (e. g., Berente et al., 2011; Bowman, Sowndararajan, Ragan, & Kopper, 2009; Dannecker et al., 2008; Kiili, 2005). Our objective with regard to virtual worlds is to

CONTRIBUTE TO A BETTER UNDERSTANDING OF INDIVIDUAL PERFORMANCE IN VIRTUAL WORLDS.

In order to identify predictors of virtual world performance, we first investigate what we refer to as inter-reality (IR), that is, the experience in-between the real and the virtual which is specific and inherent to virtual world usage. We have introduced the concept of IR previously (Schiele et al., 2011). It incorporates essential insights on the bidirectional link between the virtual and the real world (we further elucidate this concept in the remainder of this thesis). Because first and foremost, "virtual worlds are *places*" (Bartle, 2004, p. 475, emphasis in original; cf. also Goel et al., 2011), spatiality plays a key role when exploring virtual worlds (Animesh et al., 2011). Moreover, when the real and the virtual world converge (cf. Bainbridge, 2010), a new state of is created, and users find themselves in IR. Such an IR experience is potentially reinforced through one of the distinctive features of virtual worlds (cf. Barnett & Coulson, 2010): When entering a virtual world, a user is inherently disembodied (Schultze, 2010) and virtually represented by a computer-mediated character or so-called avatar that the user customizes and controls. It has previously been suggested that affordances related to this experience appeal to individual differences, as for instance differences with regard to the "cognitive and emotional relationships between a human user and his or her online representation" (Bainbridge, 2007, p. 475) as well as with regard to an IR-specific media literacy (cf. Caperton, 2010). Users' strong identification with their virtual counterpart can lead to a temporary shift of self-perception that enhances the feeling of being in-between two worlds (Klimmt, Hefner, & Vorderer, 2009). Dominant themes in virtual world thus revolve around identity, presence, immersion, and flow (Cahalane, Feller, & Finnegan, 2012; Jennett et al., 2008; Schultze & Orlikowski, 2010). Consequently, in order to develop effective and well-received systems that involve virtual experiences, human factors–besides all technological challenges–are believed to be key issues (cf. Bainbridge, 2007; Stanney, Mourant, & Kennedy, 1998; Zhao, 2011). Yet instead of taking “the nature of human and non-human interaction” into account (Cahalane et al., 2012, p. 11), current theories often take a perspective that draws a clear line between technology and human actors (cf. Schultze & Orlikowski, 2010).

In line with our literature review, we claim that well-motivated theories which sufficiently deal with the specific affordances of virtual worlds for humans are lacking (Schultze & Orlikowski, 2010; Schiele et al., 2011). We postulate that accounting for individual differences can give valuable clues about the connections between human factors and the complex mechanisms of IR, especially with regard to performance of virtual world users. Building on findings from previous research, we propose a model which links IR emotional skills–enhanced by cognitive absorption, cognitive ability, and–as an approximation of IR media literacy–experiences with IR media during childhood–to IR performance. We thereby not only draw on studies conducted in the context of virtual worlds (cf. Holsapple & Wu, 2007; Suh et al., 2011) as well as on studies relating emotional variables to IS use (e. g., Beaudry & Pinsonneault, 2010), but also on literature investigating the mechanisms of emotions, cognitive ability, cognitive absorption, presence, individual performance, the strategic use of psychological skills in sports, and media literacy (e. g., Joseph & Newman, 2010; Sacau, Laarni, & Hartmann, 2008; Jackson, Thomas, Marsh, & Smethurst, 2001; Christ, 2004).

1.2 SUBJECT AREA II: MISSING DATA

1.2.1 *Motivation*

Before being able to test our hypotheses, we were faced with a variety of challenges related to the characteristics of our data. First, our sample contained several thousand respondents, a circumstance which significantly complicates plausibility checks and preliminary analysis. Moreover, sample size impacts many evaluation indexes which serve to adjudge certain data properties or the fit of a research model. Second, during data screening for the purpose of detecting peculiarities (cf. Malone & Lubansky, 2012, p. 276), we identified response patterns in the survey data that seemed inexplicable and generated unexpected results. Third, missing values analysis unveiled a large amount of missing data, a condition under which testing whether certain relationships among data exist becomes a particularly complex task. As an example, if missing values are involved when interactions are assumed to hold between variables in a regression analysis model (cf. Baron & Kenny, 1986), the assumed interactions need to be reflected in the so-called imputation model used to treat the missings (cf. e. g., Graham, 2012a; Schafer, 2003; Black, Harel, & McCoach, 2011), which implicates that for every relationship additionally added to a research model, the missing data treatment needs to be adjusted. And last but not least, our data deviated from a multivariate normal distribution. As this is very common with real-life studies, this was not surprising; however, it emerged that the presence of missing values affects methods which are designated to account for data that deviate from normal distributions, hence distribution issues came into focus, too.

It turned out that the unexpected response patterns mentioned above had occurred due to unfortunate database operations which were performed by our cooperation partner before placing the data at our disposal. Once the source of the database errors was discovered, a re-export from the survey system remedied all of them at once and enabled us to begin afresh with the original, unmodified information. However, problems related to sample size, missing values, and data distributions did not vanish with the re-exported data. As a matter of fact, we needed to evaluate the data situation over and over and develop a specific treatment strategy at each step throughout the entire investigation process. Many technical difficulties we had to cope with during analysis were unforeseeable. For example, sometimes output data of one tool would not meet the requirements of another (even contradictory to the respective user manual), so that we needed to implement intermediate procedures to parse, process, and finally output the data in a readable format.

Examples of studies which explicitly address missing values or other data issues are scarce, although literature regarding these topics is abundant. We thus chose not only to document the different activities of hypotheses development, data-gathering, and hypothesis testing related to this research project, but also to point out the obstacles we encountered during our investigations–in particular those related to missing values–and explain the reasoning behind our decisions on how to overcome them.

This study will be of value by sharing our findings and practical experiences and will hopefully make a contribution to the community's knowledge base for adequate treatment of missing values available.[4]

4 In order not to impede the flow of reading, unless of immediate importance to comprehend the current step of analysis, aforementioned aspects are usually treated as part of the discussion.

1.2.2 *Research Objective*

Despite their omnipresence in empirical science (e. g., Nakagawa & Freckleton, 2008; Sterne et al., 2009; Paddock, 2002; Enders, 2001b; Andridge & Little, 2010; Grittner, Gmel, Ripatti, Bloomfield, & Wicki, 2011), including relevant information about missing values and informing about their treatment is far from being common with empirical studies (e. g., Roth, 1994; McKnight et al., 2007; Myers, 2011). The subject of missing data has also largely been neglected by IS literature to this day; as for instance, while "not test statistical model assumptions", "not report violations of a statistical model's assumptions", and "randomly duplicate data to increase sample size" were explicitly seen as questionable research-related behaviors in an investigation among IS researchers (Allen, Ball, & Smith, 2011), there was no mention of behavior related to missing observations. We can only conclude that awareness for threats to reliability of study results imposed by an unsuitable treatment of missing data is low in the IS community (see Schafer & Graham, 2002, for examples). Our claim is that there is a need for actively establishing a standard practice with regard to missing values among IS authors, reviewers, and editors. Our objective related thereto is to

CREATE AWARENESS FOR MISSING DATA IN ORDER TO PROMOTE STANDARDS FOR THEIR TREATMENT.

As a matter of fact, a large body of literature on state-of-the-art methods to handle missings (e. g., Collins, Schafer, & Kam, 2001; Enders, 2003) and guidelines on how to report them (e. g., McDonald & Ho, 2002; Hoyle & Isherwood, 2013) are already available from neighboring disciplines–which have traditionally taken a pioneering role in advancing statistical research and methodology. It is our belief that adapting our reporting culture accordingly will not only help to reveal the prevalence of missing data across studies and promote a "destigmatization" of the subject. Moreover, this may foster the development of new missing data techniques (MDTs), approaches, and tools which are better and more comprehensive than those currently available, and ultimately support the IS discipline's efforts to raise scientific standards and increase research rigor (cf. e. g., Boudreau, Gefen, & Straub, 2001). One objective of this thesis is therefore to create awareness for challenges related to missing data and point to available MDTs, thus hopefully stimulating the debate on extending IS reporting standards in a constructive manner (cf. Grover, Straub, & Galluch, 2009). Our attempt to achieve this aim lies in stating the various and complex data issues we encountered, carefully documenting our rational behind our decisions, and explaining how we addressed them.

1.3 RESEARCH DESIGN AND STRUCTURE OF THIS WORK

At the beginning of this research project, we needed a more accurate understanding of the concepts at the core of our research, that is, emotions, cognitive ability, cognitive absorption, and performance (see above; cf. also our research-in-progress paper, Schiele et al., 2011). For the purpose of finding suitable definitions of these constructs, we initially performed a literature review in the IS context. We then explored literature from adjacent fields for a more comprehensive view; by doing so, we followed recommendations to borrow from other theoretical fields and to adapt similar constructs when suitable theories are lacking (cf. Weiber & Mühlhaus, 2010). The results of these investigations can be summarized in the following way:

PSYCHOLOGY gave needed insights into emotional capabilities and childhood media experience, both key individual differences integrated in our model; also, we were provided with literature on (cognitive) ability and performance measures as well as on self-estimated measures;

HUMAN-COMPUTER INTERACTION (HCI) helped us to become familiar with the complex processes of immersion and its impact on humans;

DIGITAL GAMING literature introduced us to attitudes and needs of virtual world users; and finally,

(NONDIGITAL) SPORT served to approach issues related to eSports like competition, for example, as literature on eSports as such is very rare.

This enabled us to narrow down the domains of our main constructs–referred to as the primary constructs hereinafter–now also including childhood media experience, the latter of which is somewhat peripheral to the IS literature. As previously mentioned, our research area represents a relatively new field which offers limited (theoretically founded) findings to build upon. Hence our work was to a large degree of rather exploratory nature (e. g., W. W. Chin, 1998; MacCallum, Roznowski, & Necowitz, 1992), especially with regard to our hypothesized model and the characteristics of postulated effects (e. g., direct vs. indirect) on IR performance. According to our literature review in both the core of IS research and in adjacent research fields, further factors also had the potential to affect performance in virtual worlds (see also our research-in-progress paper, Schiele et al., 2011), either in the form of direct effects, or as moderators or mediators (cf. Baron & Kenny, 1986; Li & Beretvas, 2013). We therefore did not limit our investigation to the above-mentioned constructs a priori, but instead also accounted for additional constructs which were not part of the actual research model.[5] These constructs–referred to as the secondary constructs in the following–comprised self-motivational traits, competitiveness, enjoyment, parental control, and mediation. Though of lower priority to this particular study, we nonetheless explored research questions related to these constructs and further control variables as well. Moreover, we developed ideas, hypotheses, and measures involving them and reflected on according indicator-construct relationships. The task of analyzing the potential of *all* of these constructs and variables as a cause, antecedent event, or necessary condition of IR performance, however, goes beyond the present study, and an ad hoc approach additionally bears the risk of capitalizing on chance (cf. W. W. Chin, 1998; W. W. Chin & Todd, 1995). Instead of optimizing our initial model according to statistics (MacCallum et al., 1992), we thus concentrated on testing our initial hypotheses. The aim of this thesis is to provide a complete picture of our venture, with particular focus on all kinds of data manipulations applied, the methods of analysis used, and possible limitations. We also attempt to present the domain conceptualizations, related hypotheses, and corresponding measures of *all* constructs in sufficient depth. With regard to the resources available, we yet only relate to our secondary hypotheses and control variables where necessary, and only to the extent needed for a better understanding. Additionally incorporating more variables could certainly serve to complement the results of this study in the future, but this task exceeds the scope of the present study.

The next steps consisted of defining the conceptualizations (cf. Suddaby, 2011) of our constructs, developing hypotheses about their casual structure and their predictive power with regard to IR performance, depicting these hypotheses with the aid of a graphical model (Bagozzi & Yi, 2012), and determining the assumed type (reflective vs. formative, see, e. g., Bollen, 2011) of relationships between the latent variables and their indicators (herein after referred to as factor-indicator relationships) for each construct. In line with the type of research questions we were interested in (cf. Weiber & Mühlhaus, 2010, p. 19), that is, research questions which impose the testing of a hypothesized model of relationships between latent, nonobservable variables (also termed constructs, concepts, etc., see Bagozzi & Yi, 2012), our study was designed to suit the application of structural equation modeling (SEM); advantages of this method of analysis for this type of research are well-elaborated (e. g., Bollen & Hoyle, 2012; Gefen, Rigdon, & Straub, 2011; Animesh et al., 2011; Cenfetelli & Bassellier, 2009; W. W. Chin, 1998; Baron & Kenny, 1986; see also Hoyle, 2012b; Bagozzi & Yi, 2012; Green & Thompson, 2012, for a comparison with other methods of analysis).

5 See the *Publication manual of the American Psychological Association*, 2011, p. 28 and p. 35, on expounding the prioritization of research goals and distinguishing between primary and secondary hypotheses.

Subsequently, appropriate instruments to measure our variables of interest were required. In most cases, these were operationalized in the form of item-based measures, while others consisted of single data point measures. On the basis of our extensive literature review across several research disciplines, we generated candidate sample items for all constructs, mostly from existing scales, and assessed their validity with the help of members of the community. Furthermore, we prepared additional variables for possible multigroup comparisons and control purposes. We then needed to collect actual data to test our model on. To this end, we cooperated with the operator of one of the largest e-Sports communities in Germany, a circumstance which permitted us to collect survey and performance data amongst its virtual world users. To first evaluate and pretest the items before the actual study (see, e. g., Straub, 1989; and Boudreau et al., 2001), we conducted three pilot studies of different sizes–hereinafter referred to as Pilot 1, Pilot 2 and Pilot 3–with three different subsamples of the community. After altering and eliminating inadequate candidate items according to the pretest results, we proceeded with an adjusted item list and conducted the actual survey with a new sample. Our final questionnaire was then included in the community's regular survey. Said survey is conducted by the community portal operator and sent to all registered members usually every two years. Approximately 5,500 survey candidates responded in total, not accounting for missings. In addition, we collected members' performance data with the aid of a web crawler. The two sources of data, questionnaire data and web crawler data, were then merged in order to be able to analyze the hypothesized antecedents to performance and their impact on performance simultaneously. Next, we needed to test our hypothesized model with the data collected and interpret the analysis results.

At all stages of the study, we were guided by a number of state-of-the-art publications including, but not limited to:

- the measurement development procedure proposed by MacKenzie, Podsakoff, and Podsakoff (2011), an update of "Churchill's (1979) seminal article" (MacKenzie et al., 2011, p. 294) for reflective measures (or scales), and for formative measures (or on scales and indexes) by Diamantopoulos and Winklhofer (2001); further

- the criteria for correct measurement model specification developed by Jarvis, MacKenzie, and Podsakoff (2003) and revisited by Petter, Straub, and Rai (2007),

- the SEM guidelines published by Bagozzi and Yi (2012), Hoyle (2012a), Kline (2011), Vinzi, Chin, Henseler, and Wang (2010), and Wetzels, Odekerken-Schröder, and van Oppen (2009)–that is, with regard to covariance-based structural equation modeling (CBSEM) and variance-based SEM/partial least squares (PLS), respectively (cf. Haenlein & Kaplan, 2004),

- the view of Schafer and Graham (2002) on missing data and the recommendations of the Task Force on Statistical Inference of the American Psychological Association (APA) for documenting missing data in empirical studies (as cited and extended by McKnight et al., 2007),
- the reporting and normative standards required by Allen et al. (2011), Ringle, Sarstedt, and Straub (2012), and Gefen et al. (2011),

and many more. We explain which type of procedure we finally applied during the study and our rationale related thereto in the upcoming chapters.

The remainder of this work is structured as follows: The second chapter elaborates on the theoretical background and conceptual framework of our virtual world study and states our research hypotheses. The third chapter treats the development of measurements and describes the conduction of the pretests as well as of the actual study. The fourth chapter explains the data preparation, details the methods used to analyze our data at different stages, and reports the analysis outcomes. The fifth chapter interprets our results and discusses possible limitations. Finally, the sixth chapter evaluates the contributions of this study and concludes with avenues for future research.

2 PRIOR RESEARCH, MODEL, AND STUDY HYPOTHESES

Essentially all models are wrong, but some are useful.

— *George E. P. Box*

In the following, we first outline our previous work on the topic of this study. Subsequently, we elucidate the theoretical foundations of our research and depict our research model. We then develop our hypotheses on predictors of IR performance and present the definitions of our constructs.

2.1 PREVIOUS WORK AND THE PRESENT STUDY

As expounded in the introduction, parts of this chapter build on a research-in-progress paper which we submitted at a preliminary stage of our research and presented at ICIS 2011 in Shanghai, China (Schiele et al., 2011); the model of IR literacy that we proposed in the aforementioned publication is depicted in Figure 1 and is the predecessor of the model in the present study.[1] The previous model, too, drew on the meta-analysis conducted by Joseph and Newman (2010), who investigated the relationships between job performance, cognitive ability, and the three emotional intelligence (EI) subfacets emotion perception, emotion understanding, and emotion regulation. The latter tested their conceptual model by constructing a correlation matrix from estimates extracted from other meta-analyses. In an additional post hoc analysis, they also accounted for a possible moderating effect of a job's demands for emotional labor, so for the level of emotion regulation in terms of feelings and expressions required to successfully accomplish a job (results are represented in Figure 3).

Rather than just reproducing the original exposition of our theoretical arguments presented at ICIS 2011, in this thesis, we elaborate on findings from our literature review in much more detail, aiming at better specifying the rationale of our research approach and incorporating some of the feedback we received concerning our publication.

1 Note that the present study uses a slightly different model. For example, we called the central construct of our model IR emotion handling (IEH). However, it turned out that this term was chosen rather unfortunately, because it can be confused with a possibly unusual behavior that may not be representative for a person, but rather exhibited in an exceptional situation and of purely momentary nature. We therefore returned to the original names of the EI facets which frame our understanding of emotions in this study, to better reflect the fact that these–acquired or innate–abilities represent individual differences and are inherent to a person.

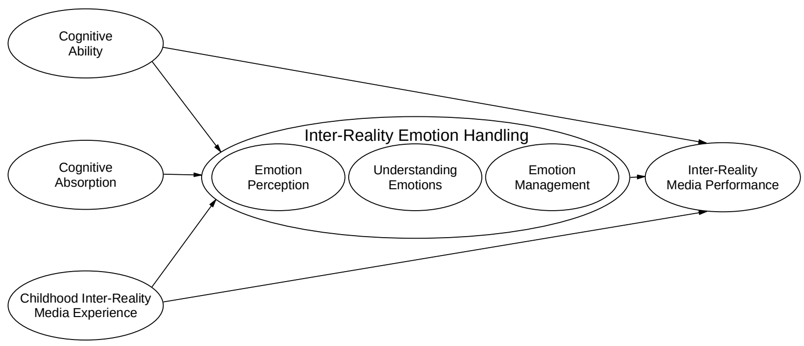


Figure 1: Our original model of IR emotion handling (IEH) as presented at ICIS 2011 during the research-in-progress session; all arrows represent assumed positive effects regarding the hypothesized latent variables involved. Adapted from: Schiele, S., Weiss, T., & Putzke, J. (2011). On Inter-Reality Literacy: Emotions as Predictors of Performance in Virtual Worlds [Research in progress]. In D. F. Galletta & T.-P. Liang (Eds.), Proceedings of the Thirty-Second International Conference on Information Systems.

Moreover, since the presentation of our research proposal, we also had the chance to broaden our knowledge through another study in the fields of virtual worlds which we also conducted around that time. Aforementioned study indicated that competition, challenge, and escapism are strong drivers for continuous virtual world use in the context of top-league eSports (Weiss & Schiele, 2013). Additionally, we further extended our knowledge by searching news and game forums for related content and by interviewing virtual world users one-on-one, the ultimate goal being to strengthen the theoretical foundations of our research approach. Meanwhile, we also needed to acknowledge many articles on virtual worlds which have been published between the ICIS 2011 submission and today. *MIS Quarterly*, for example, dedicated a special issue to this subject area in September 2011 (call for papers: Jarvenpaa et al., 2007). Altogether, our supplementary research activities lead on to a refinement of our theoretical and methodological considerations. As a result, we revised, completed, and extended some of our previous propositions and reflected these adaptations through a more comprehensive framework. At last, we conducted this study in its present form. We go into more details below.

As mentioned earlier, we distinguish between primary and secondary constructs and hypotheses.[2] Hence, to increase readability during the development of our model propositions, we introduce concepts, constructs, hypotheses as well as analysis activities accordingly. In this vein, the primary constructs frame the initial model, and hypotheses–referred to as our primary hypotheses–represent suggested relationships

2 See also *Publication manual of the American Psychological Association*, 2011, p. 28 on the distinction between primary and secondary hypotheses.

between primary constructs. Furthermore, we suggest that additional constructs and variables, namely our secondary constructs as well as certain controls, should also be accounted for. Said constructs and variables were also included in the survey and are presented in the course of this thesis as well. Related hypotheses–referred to as our secondary hypotheses–postulate certain direct or indirect effect (cf. e. g., Kline, 2011; Marsh, Wen, Nagengast, & Hau, 2012; Jaccard & Wan, 1995) of secondary constructs on primary constructs (e. g., moderators or mediators additional to the initial model, cf. Bagozzi & Yi, 2012, p. 30) and propose that they shed light on group differences in our sample (see Millsap & Olivera-Aguilar, 2012; Byrne, 2004; Raykov, 2004, on measurement invariance). However, restrictions on time and space did not allow to test our secondary hypotheses except one, namely h17 which relates to the moderator effect of emotional labor suggested by Joseph and Newman (2010). By applying a distinction between primary and secondary hypotheses where necessary, we aim at indicating the importance and respective contribution of a particular analysis as well as related activities with respect to the study as a whole.

2.2 VIRTUAL WORLDS

A large variety of technologies which create virtual experiences exists nowadays. While some environments build a reality of their own which one can enter and leave, like for instance, hybrid environments with visual and haptic feedback or computer-assisted virtual environments known as CAVEs (cf. Schroeder, 2006; Rosenbloom, 2004), others involve augmenting or alternating the actual[3] reality through additional artificial features that can be switched on and off (Nieuwdorp, 2007). Such experiences can be supported by means of head-mounted displays (Vasilakos et al., 2008), immersive projection technology (Schroeder, 2002), or mobile phones–such as, for example, with Ingress, a pervasive (cf. Nieuwdorp, 2007) massively multiplayer online game (MMOG) where players which are equipped with mobile phones are requested to stay in physical proximity to real-world sites in order to interact with the sites' virtual representations.[4] Furthermore, new haptic devices are being developed to supply future collaborative virtual environments with emotional warmth, nonverbal intimacy, and more cues to express and recognize emotions, in order to enhance the realism of interactive experiences (Bailenson, Yee, Brave, Merget, & Koslow, 2007; Basori, Daman, Sunar, & Bade, 2007).

Virtual worlds, however, "contrast strongly with the concept of totally immersive virtual reality (VR)" .... Images are restricted to the

3 As Schultze (2010) remarks, the term "real" as an antonym for "virtual" is problematic. She proposes to apply the term "actual" where suitable; see also Cahalane et al. (2012) concerning this conception.

4 http://www.ingress.com/

screen of an ordinary computer monitor, rather than filling the walls of a VR cave or a binocular head-mounted display" (Bainbridge, 2007, p. 475). Additionally, virtual worlds' sensory perceptions are currently reduced to "sight and sound" and thus exclude "smell, taste, and touch" (Wasko et al., 2011, p. 649). On the other hand, different than typical virtual reality (VR) environments, virtual worlds allow for multi-user participation and are not limited to short sessions (Schultze, 2010). Attempts to narrowly define virtual worlds often combine descriptions of their technological features with descriptions of activities they allow for. Correspondingly, virtual worlds offer "real-time communication, freedom to navigate and manipulate objects, and interactivity among users" (Nah et al., 2011, p. 732). Being "multi-modal platforms that feature rich graphics, 3D rendering, high-fidelity audio (...), video, [and] motion" (Schultze, 2010, p. 434), they are often said to provide richer experiences than other technologies (Wasko et al., 2011; Barnett & Coulson, 2010; Y. Lee, Kozar, & Larsen, 2009). In their entirety, virtual worlds are characterized as electronic, computer-generated, synthetic, web-based, persistent, and immersive environments (Animesh et al., 2011; Bainbridge, 2010; Goel et al., 2011; Castronova, 2005; J. Mueller, Hutter, Fueller, & Matzler, 2011). They are regarded as "utterly contrived and artificial" (Schultze, 2011, p. 2) "alternative versions of the real world" (Suh et al., 2011, p. 711) or "virtual alternate realities" (Chaturvedi et al., 2011, p. 674) which "visually mimic (...) complex physical spaces" (Bainbridge, 2007, p. 472).

Nevertheless, virtual worlds also "mimic the real world in which players form friendships and communities" (Barnett & Coulson, 2010, p. 167). Users of virtual worlds can interact with other computer-generated individuals, objects, and landscapes (Bray & Konsynski, 2007; Nah et al., 2011) "in a realistic manner" (J. Mueller et al., 2011, p. 1), such that this interaction is "comparable" (Bainbridge, 2010, p. 1) to that of the real world. As a matter of fact, virtual worlds allow to "replicate almost any real-world activity. That includes conversations, battles, sex, and, increasingly, commerce" (Castronova, 2005, p. 20). Participants of virtual worlds can thus not only communicate synchronously (Goel et al., 2011), live "versatile virtual lives" (Partala, 2011, p. 787), and engage in civic participation (Bers, 2010), but also "virtually run global businesses in real-time" (Bray & Konsynski, 2007) and take part in economies (Annetta, 2010). Commercial activities connected to virtual worlds already involve "a wide range of virtual products and services (Animesh et al., 2011, p. 790), marketing (Nah et al., 2011) as well as the exchange of money (Messinger et al., 2009), for instance in the form of shopping (Animesh et al., 2011). Researchers have coined the term virtual commerce to subsume this type of business (Arakji & Lang, 2008).

To capture and better understand virtual worlds and their relation with human cognition and emotions, one needs to further examine its distinguishing features related to corporeality and spatiality (cf. Schultze & Orlikowski, 2010; Barnett & Coulson, 2010). The following section deals with the concepts of avatars and IR in more detail. Aspects of virtual worlds related to location are discussed further below.

2.2.1 *Characteristic Features*

As the virtual and the real world mutually influence each other in the context of virtual worlds, their boundaries become blurry (cf. Suh et al., 2011; Schiele et al., 2011). Evidence for a transition from a state of dual reality (where reality and virtual reality interact but still coexist) to a state of mixed reality (where both realities "exhibit highly correlated stable, phase-locked periodic motion") can be found by means of experiments in physics, such that a real pendulum and a virtual pendulum move together as one (Gintautas & Hübler, 2007, p. 2). Similarly, interactive virtual environments also attempt to create a state of mixed reality, notably by combining both virtual and real world environments and physically incorporating the people wishing to experience this environment (Schaik, Turnbull, Wersch, & Drummond, 2004). Users of virtual worlds, however, are not "in" the virtual world themselves, but remote (Clayes & Anderson, 2007). As such, they can be described as disconnected from their bodies or disembodied by computer-mediation (Bray & Konsynski, 2007; Schiele et al., 2011; Schultze, 2010). Instead, virtual worlds allow their users to create computer-animated characters (Bray & Konsynski, 2007), so-called avatars. Such avatars are digital representations of themselves and provide them with a simulated body which can move around in the virtual environment (Nah et al., 2011; Petrakou, 2010). The term avatar originates from Sanskrit and can be translated to incarnation (Messinger et al., 2009; Y. Lee et al., 2009). Virtual worlds are thus environments in which users participate (Messinger et al., 2009), that is, "inhabit, socialize, and perform economic activities" (Animesh et al., 2011, p. 790), with the aid of avatars as their virtual incarnation or graphical embodiments (Y. Lee et al., 2009; Goel et al., 2011). Some therefore consider avatars to be mental representations that are utilized to include oneself in the external *and* the virtual world, a process termed distal attribution (Nash, Edwards, Thompson, & Barfield, 2000).

Literature suggests that the representation through avatars and the resulting interaction between individuals, virtual objects as well as the virtual environment is a distinctive feature of virtual worlds (Suh et al., 2011; Schultze, 2010). Owing to these and other characteristics, it has been stated that virtual worlds are something entirely new: "It's not a game, it's a... simulation. Or a service. Or a medium. Actually, it's none of these: It's a place" (Bartle, 2004, p. 473; see also Saunders, Rutkowski, Genuchten, Vogel, & Orrego, 2011). These observations give rise to the assumption that virtual worlds provide a very particular experience, inherent to their use and supported by the involvement with one's avatar(s). We refer to this experience as IR (Schiele et al., 2011).

It can be seen from the above that virtual world practices entail a variety of perspectives and roles. In order to provide clarity, we adopt the following terminology:

> The term *user* indicates a participant in a virtual world, regardless of the type of participation (e.g., player, actor, participant). A *person* is the user in the physical world, and an *avatar* represents the user within the virtual world (Wasko et al., 2011, p. 646, emphasis in original).

Being able to control their virtual representation or to have input into its design is important to users (Schroeder, 2002). Avatars can be customized with regard to various characteristics concerning physical appearance, traits, or capabilities. Though some prefer avatars to accurately reflect their offline self (Schultze, 2011), characteristics of avatars can very much differ from those of the users which control them. As an example, both the male and the female sex enjoy "gender swapping", that is, male users choosing a female avatar and vice versa (Hussain & Griffiths, 2009). Also, avatars may resemble humans, but depending on the theme or content of a virtual world (Messinger et al., 2009), they may look like monsters, elves, dwarfs, or gnomes. Some avatars are able to run very fast, climb walls, or fly, and while some are violent and aggressive, others have the ability to heal or perform various forms of magic (Schiele et al., 2011; Ducheneaut, Yee, Nickell, & Moore, 2006).

Through the process of re-embodiment, users are enabled to act out new roles and build a new identity or image of themselves (Frostling-Henningsson, 2009), for example, that of a hero or villain (Barnett & Coulson, 2010). By virtually adopting the avatar's characteristics, they feel like–or even become–the other "person" (Zhou, Jin, Vogel, Guo, & Chen, 2010). This kind of identification has been defined as "the cognitive connection between an individual and an avatar, with the result being that the individual regards the avatar as a substitute self or has such an illusion" (Suh et al., 2011, p. 715). Such a state in which users experience their virtual self as if they were their actual self is not limited to avatar representation but can also occur when one's body is represented through a mediated version of oneself–for example in a video conferencing system. In both situations, users do not notice the virtuality (Park, Min Lee, Annie Jin, & Kang, 2010). It has been observed that avatar identification can lead to such a strong entanglement between users and their avatars that a temporary shift in self-perception can occur (Schultze, 2011; Klimmt, Hefner, & Vorderer, 2009). Different types of avatar-self relationships related to virtual worlds have been identified (Schultze & Leahy, 2009).

Besides giving the user a body in virtual worlds, avatars further supply him or her with what is referred to as presence, an essential concept related to immersive technologies that fosters an IR experience (Schultze, 2010; Schiele et al., 2011). Presence harks back to telecommunication psychology and has often been applied to capture and describe properties of media (or IS). Many perspectives on presence exist, and various attempts have been made to differentiate the many conceptualizations. We debate this subject area below when taking a media perspective on virtual worlds.

2.2.2 *Perspectives and Applications*

Some refer to virtual worlds as new media (Berente et al., 2011; cf. also Bracken & Skalski, 2009), whereas others argue that virtual worlds are not a medium in terms of a communication channel themselves but rather incorporate means for communicating (Bartle, 2004). Again others claim that virtual worlds represent a new type of IS (Chaturvedi et al., 2011), while some think of virtual worlds as built on top of an IS (Wasko et al., 2011). This indicates that different contexts of virtual world usage and their actual application call for different perspectives and levels of abstraction.

VIRTUAL WORLDS AS AN IS Taking an IS view, virtual world studies can draw upon research on individual technology adoption and on the impact of IT artifacts (cf. Lucas, Burton Swanson, & Zmud, 2007; Hirschheim, 2007; Venkatesh, Davis, & Morris, 2007). One of the underlying assumptions of this research stream is that only an IT which is actually being used may have a positive impact on individual performance (cf. Goodhue & Thompson, 1995). The most prominent model in this regard, the Technology Acceptance Model (TAM), is considered the "most influential and commonly employed theory in information systems" (Benbasat & Barki, 2007). The model itself is based on a body of research on belief, attitude, intentions, and behavior and attempts to predict and explain IS use with the aid of two variables, namely perceived usefulness and perceived ease of use (Goodhue, 2007; F. D. Davis, 1989). Subsequent research has re-validated the model (Adams, Nelson, & Todd, 1992; Segars & Grover, 1993) and confirmed its hypothesized effects "almost to the point of certainty" (Benbasat & Barki, 2007, p. 212).

Nonetheless, studies on usage caution to expect consistent effects related to a particular technology. Many individual differences, notably age, gender, experience, personality traits, cognitive style, expectations, or type of training–to name just a few–have been found to affect usage, satisfaction, and ultimately the effectiveness of system usage (e. g., Adams et al., 1992; W. W. Chin & Lee, 2000; Devaraj, Easley, & Crant, 2008; H.-W. Kim & Kankanhalli, 2009; Venkatesh, L. Thong, & Xu, 2012). Moreover, empirical evidence suggests that an individual's usage of, for example, the Internet or social networks may be influenced by the usage patterns exhibited by that individual's peers–among children and teenagers as well as adults (Agarwal, Animesh, & Prasad, 2009; Sykes, Venkatesh, & Gosain, 2009; Hundley & Shyles, 2010). Similarly, through complicated mechanisms, online user reviews of software products have an impact on sales (Duan, Gu, & Whinston, 2009). Literature has thus extended the TAM many times and adapted it to various application areas (Devaraj et al., 2008; cf. Venkatesh, 2000; Venkatesh, Morris, Davis, & Davis, 2003, 2012; S. Kim, 2009; Ha & Stoel, 2009), yielding a great number of independent variables for predicting intentions and behavior (Bagozzi, 2007).

Overall, this resulted in "a state of theoretical confusion and chaos" (Benbasat & Barki, 2007, p. 212), and some have called the scientific substantiation of findings related to this stream of research into question (Straub & Burton-Jones, 2007; Silva, 2007). General issues of the approach relate to the fact that utilization is often not voluntary, that even extensive use or the use of a poor IS will not (automatically) improve performance, and that "poor systems may be utilized extensively due to social factors, habit, ignorance, availability, etc., even when utilization is voluntary" (Goodhue & Thompson, 1995, p. 216). The construct of IS use is still weakly conceptualized and operationalized as "frequency, duration, or variety of system functions used", hence does not take into account the multidimensional nature of IS (Barki et al., 2007, p. 173).[5] Furthermore, IT acceptance has often been interpreted in terms of the narrow notion of IS use (A. Schwarz & Chin, 2007).

In response to this situation, researchers have attempted to shed light on the "multiple roles that people fulfill while adopting, adapting, and using information systems" (Lamb & Kling, 2003, p. 197). This has also given rise to intensive research on habits and other post-adoption issues related to social norms or emotions, for instance (Limayem, Hirt, & Cheung, 2008; Ortiz de Guinea & Markus, 2009; Polites & Karahanna, 2012; Beaudry & Pinsonneault, 2010), often with a focus on discontinuance on the organizational level (Silva & Hirshheim, 2007; Furneaux & Wade, 2011). Additionally, different measures of satisfaction, information needs, performance expectancy, and other requirements related to productivity were developed (W. W. Chin & Lee, 2000; W. W. Chin & Marcolin, 2001; Bélanger, Collins, & Cheney, 2001; Venkatesh et al., 2003).

Nonetheless, attempting to conceptualize virtual worlds as a form of IS calls for additional research. For example, virtual worlds allow for the pairing of a virtual system and its real-world counterpart and to function as an IR system (cf. Gintautas & Hübler, 2007). To achieve

5 See Burton-Jones and Straub (2006) for an overview of a categorization of measures of usage ordered from lean to rich, where lean measures reflect use or nonuse alone, and rich measures reflect the nature of usage, "involving the system, user, and/or task", p. 233.

such a coupling, a company wishing to use this technology has to put a certain effort into bridging the gap between real world and virtual world services, for example, through adjustment of connected work flows (Kadavasal, Dhara, Wu, & Krishnaswamy, 2007). Challenges related to such an approach may be embedded in the wider context of process virtualization, that is, processes that incorporate "physical interaction between people and/or objects which are transitioned to a virtual environment" (Overby, Slaughter, & Konsynski, 2010, p. 700). Furthermore, virtual worlds differ from solely productivity-oriented IS in that they may serve both utilitarian and/or hedonic purposes. This circumstance has implications for their use and adoption (Wakefield & Whitten, 2006; Holsapple & Wu, 2007; Hsu & Lu, 2004; Mantymaki & Riemer, 2011): Not surprisingly, the predictive value of "fun-aspects" related to IS use (like ease of use and enjoyment) outweighs that of perceived usefulness in hedonic contexts (Heijden, 2004). This may be particularly advantageous considering that play, playfulness and cognitive absorption, thus constructs linked to virtual world usage, are important aspects of work and influence beliefs about the usefulness or the ease of use of the system (Agarwal & Karahanna, 2000; Goel et al., 2011; Wakefield & Whitten, 2006; Webster & Martocchio, 1992). On the other hand, the hedonic benefits may rather foster prolonged than productive use (Heijden, 2004).

Another aspect that is essential to virtual worlds is their emergent nature: While former IS are considered inflexible and unsuitable to adapt to a constantly changing environment, virtual worlds "combine the structural aspects of traditional modeling and simulation systems in concert with emergent user dynamics of systems supporting emergent knowledge processes" (Chaturvedi et al., 2011, p. 673; cf. also Truex, Baskerville, & Klein, 1999; Cahalane et al., 2012). It is also noteworthy that gender seems to moderate how perceived benefits–hedonic, utilitarian, and social–of using virtual worlds impact satisfaction and continuance intention with regard to the virtual world usage experience: While for male users, the hedonic benefits associated with virtual worlds plays a dominant role in promoting their satisfaction, all three types of benefits are almost equally important to female users (Zhou, Jin, & Fang, 2014). Also, users who regularly visit virtual worlds may encounter different issues than those with little virtual world experience, or those with computer anxiety (Animesh et al., 2011; Webster & Martocchio, 1992; Jarvenpaa & Staples, 2000).

VIRTUAL WORLDS AS A MEDIUM Taking a media perspective, knowledge from the research field of new media are applicable to virtual worlds. Owing to many shared characteristics, virtual worlds are often considered new media themselves (cf. Leung & Lee, 2005; Berente et al., 2011; Bracken & Skalski, 2009). For instance, virtual worlds qualify as highly interactive media which increase connectivity and underlie effects of network externalities (cf. H. Liu, 2010). Hence perceived enjoyment and social norms should be important for their acceptance (cf. Dickinger, Arami, & Meyer, 2008). Also, as indicated by studies on other Internet-based technologies such as blogs, purpose and operation of virtual worlds as well as their effects on communication are likely to play an important role in this respect, too (cf. Hsu & Lu, 2004). Communication, collaboration, and cooperation, summarized as community factors, have previously been identified to particularly influence acceptance of virtual worlds and related use intentions (Fetscherin & Lattemann, 2008). To further investigate these aspects and to evaluate "how these media (...) fit into our everyday lives" (Schroeder, 2002, p. 10), we study the social interaction they allow for–also in comparison with other means of communication–below.

Literature suggests that conceptualizing users as social actors puts researchers in a better position to

> "ask with whom an actor is interacting, about what issues, under what conditions, for what ends, with what resources, etc. It is a metaphor that readily expands the scope and scale of the social space of people's interactions with ICTs and with other people, groups, and organizations" (Lamb & Kling, 2003, p. 224).

This approach particularly provides opportunities for advancing our understanding of virtual worlds' communication effects. As aforementioned, one of the distinct features of virtual worlds, that is, the re-embodiment through avatars, allows a user to "be" in the virtual world. This experience resembles a phenomenon rooted in media research which is referred to, among others, as *presence*. As Biocca (1997) noted, the term presence first and foremost captures the "perceptual sensation of being in a place other than where your physical body is located" (no page number). This (as well as related phenomena) has been given many names and defined in similar ways many times, as for example in IS, communication research, human-computer interaction (HCI), and engineering. For a comprehensive summary on presence, the reader is referred to Schultze (2010) and K. M. Lee (2004). We broadly discuss various distinctions and similarities in the following.

Apart from the concept of presence, the notion of social presence, telepresence, virtual presence, spatial presence, physical presence, self-presence, and co-presence are also discussed in the context of media and virtual worlds. Some, like social presence for instance, have been conceptualized in many ways (Animesh et al., 2011). The aforementioned terms are either used interchangeably, often by reason of simplicity (cf. Bracken & Skalski, 2009), or defined narrowly, in order to correspond to special research objectives (Schubert, 2009).

Presence, for instance, has been defined by K. M. Lee (2004) as "a psychological state in which the virtuality of experience is unnoticed" (p. 32), while according to Schultze (2010), presence relates to "the user's sense that she exists in a given setting, be it virtual or actual" (p. 435). Several factors have been found to frame the concept of presence, notably the sense of physical space and engagement (Lessiter, Freeman, Keogh, & Davidoff, 2001). To the same effect, the concept of *spatial* presence refers to the sense of being "there" in virtual and real environments as well as to experiences which emerge from reading and remembering (Schubert, 2009). *Social* presence, however, has been has defined by some as the degree to which a user perceives other people to be physically present when interacting with them (Carlson & Davis, 1998). On the other hand, others have described social presence as the extent to which a medium "enables an individual to experience others as being *psychologically* present [emphasis added]" (Animesh et al., 2011, p. 792; see also Zhu, Benbasat, & Jiang, 2010). Generally, social presence theory assumes that the more social cues a medium conveys, the more it will be perceived as warm, personal and sociable (e. g., Yoo & Alavi, 2001). A study in the context of virtual worlds found that social presence was the only social outcome which had a significant impact on users' intention to use this technology (Mantymaki & Riemer, 2011). *Tele*presence, on the other hand, has been defined as a as the "experience of seeming to be present in a remote environment by means of a communication medium" (Animesh et al., 2011, p. 792) with the result of a user perceiving his or her experience in the mediated environment "as first-hand, or direct" (T. Kim & Biocca, 1997, no page number). Thus telepresence is often used when relating to remote real environments, whereas *virtual* presence is specifically applied when speaking of virtual environments (Schubert, 2009). Physical and self-presence are both physiological states. However–and somewhat confusing-Biocca's (1997) definition of a presence attributed *physical* stresses the feeling of being transported from an actual to a virtual environment and describes a state in which virtual physical objects are perceived as actual physical objects (Park et al., 2010; K. M. Lee, 2004). In contrast, *self*-presence relates to avatar-identification aspects, thus a state where "the virtual self is experienced as if it were the actual self" (Park et al., 2010, p. 824). And finally, *co*-presence lies at the intersection between tele- and social presence and refers to a sense of collocation or "the sense of being in a shared virtual setting with remote others" (Schultze, 2010, p. 438).

In the context of virtual worlds, Saunders et al. (2011) have recently focused on two interpretations of presence, namely in the form of social richness (based on social presence theory) and in the form of immersion. As indicated above, social richness appraises the perception of media according to a medium's ability to establish a personal connection through the amount of human warmth, intimacy, and sociability transmitted (Sia, Tan, & Wei, 2002; Zhu et al., 2010). Immersion qualifies the extent of perceptual and psychological immersion of a person into a virtual environment, thus "the extent to which the person seems to be immersed or engaged in the virtual world" (Saunders et al., 2011, p. 1086, citing Biocca and Levy, 1995[6]). Thanks to the "bodily practices such as sitting, gesturing, smiling, and dressing" they allow for, virtual worlds are considered "potentially more immersive than other media" (Schultze & Orlikowski, 2010, p. 812). A commonly adopted understanding of immersion is also that it refers to "a quantifiable aspect of display technology"–measurable in four dimensions, namely (a) the degree to which a user can block out distractions from the real world, (b) the number of sensory modalities the system allows for, (c) how panoramic a display is, and (d) how vivid the display is–perceived in relation to its resolution (Nash et al., 2000, p. 22; cf. also Schultze, 2010). The term of "3D immersion" has been introduced in the context of virtual worlds (Wasko et al., 2011).

Though they share a common theme, immersion and the IS construct of cognitive absorption are not congruent. The original concept of cognitive absorption accounts for (a) temporal dissociation, (b) focused immersion, (c) heightened enjoyment, (d) control, and (e) curiosity (Agarwal & Karahanna, 2000). This definition combines a personal trait (i. e., curiosity) and past experiences with a system into a general attitude towards a particular IS (cf. Jennett et al., 2008). Also, as noted by Saunders et al. (2011), Agarwal and Karahanna (2000) speak "about focused immersion and cognitive absorption as constructs related to involvement with systems in general" (p. 1086).[7] Immersion, on the other hand, solely refers to a state during a specific experience and does not account for motivations. The concept of (psychological) immersion refers to involvement and emotional engagement (Schultze, 2010). While users are interacting with an IS "there could be occasions where immersion is low" (Jennett et al., 2008, p. 643), a view which coincides with the interpretation of immersion as a multidimensional continuum (Bowman et al., 2009).

The relevance of (various forms of) presence for immersive technologies and virtual worlds in particular has been repeatedly emphasized (Schultze, 2010; Nah et al., 2011; Animesh et al., 2011; Park et al., 2010).[8] However, the question remains of how the sense of presence is generated. Even today, there is agreement with views of ancient Greece that the movement of objects is important in conceiving a place, the latter of which matters to grasping the nature of presence (Saunders et al., 2011). Some assume that the representation of action plays a pivotal

6 Biocca, F., and Levy, M. R. 1995. Communication in the Age of Virtual Reality, Hillsdale, NJ: Lawrence Erlbaum Associates (as from the references from Saunders et al., 2011, p. 1096).

7 In said conceptualization, focused immersion is actually a subdimension of the higher-order construct cognitive absorption (cf. Agarwal & Karahanna, 2000).

8 MITpress has even dedicated a whole journal of that name to the subject area: *Presence*: Teleoperators and Virtual Environments, http://www.mitpressjournals.org/pres. It addresses mechanical and electrical engineers, computer scientists, high-tech artists, media researcher and psychologists involved in the study of HCI as well as sensorimotor and cognitive behavior.

role in this regard, because it has become apparent that "the possibility to move the virtual body in the environment, to interact with virtual objects and agents, and even the mere illusion of interactivity lead to an enhanced feeling of spatial presence" (Schubert, 2009).

A crucial factor related to the representation in virtual worlds is (again) the re-embodiment of users through avatars. Presence has also been circumscribed as the extent to which users physically attribute themselves to a virtual world by means of their avatar as their mental representation (Nash et al., 2000). Through re-embodiment and avatar-identification, users not only experience their avatar as an extension "of an actual human mind translated into a virtual body", but also receive the actual feedback of "a human mind seeing oneself as a body present in a virtual world" (Bray & Konsynski, 2007, p. 21). Some definitions of presence imply "that the user of the medium considers the items in the mediated environment as unmediated and reacts directly to the items as if they are physically present objects" (T. Kim & Biocca, 1997, no page number), suggesting that an avatar is perceived as an extension of the physical body or of the user as an actor (Schultze, 2010; A. Davis et al., 2009).

Experimental psychology gives further point to this notion. To understand how, it is necessary to become familiar with the concept of object affordances. It refers to the idea that an object provides a viewer with a variety of possible actions, for example, throwing, kicking, or grasping it, or sitting or standing on it. Object affordances are "dispositions [that] arise as a result of adaptation of the nervous system over both evolutionary time scales and the lifetime of an individual" (Ellis & Tucker, 2000, pp. 162–163). Reactions on how to handle objects affordances, so-called motor programs, are therefore specific to a human through the pattern of stimulation the object provides to him or her specifically. Studies have substantiated that the activation of motor programs is automatic. Moreover, it is not only irrespective of the medium or the viewer's intentions related to the object, but also of its presentation–that is, whether it is a real object, a picture of a real object, a computer-generated rendering of a virtual object, or a remembered object (Ellis, Tucker, Symes, & Vainio, 2007; Symes, Ellis, & Tucker, 2005; Goel et al., 2011).

Consistent with these insights to the human brain, studies trying to reveal associations between the sense of presence and individual differences have naturally focused on spatial intelligence, but also on other cognitive abilities (like verbal intelligence), personality, or computer experience (Alsina-Jurnet & Gutiérrez-Maldonado, 2010; Sacau et al., 2008). In turn, virtual worlds provide opportunities to further develop sensory, perceptual, and attentional abilities which are known to be essential for tasks in spatial cognition, like spatial resolution, the attentional visual field, enumeration, multiple object tracking, as well as visuomotor coordination and speed (Spence & Feng, 2010)–however, attentional capacity, for example, does not seem to benefit from virtual world use (Irons, Remington, & McLean, 2011).

Notably, conclusions on the impact of presence on individual performance drawn by various studies differ. Some results suggest that a higher level of presence is always preferable than a lower level when intending to increase the performance of individuals, for instance for memorization tasks (e. g., Bowman et al., 2009; Ragan, 2010). In contrast, some suggest that there is no evidence for a causality between presence and in-world task performance (Sacau et al., 2008; Schultze, 2010). For example, a study found that overall, presence and performance were positively correlated, but that the relation was quite weak, and that all other factors accounted for in the study taken together were responsible for 90% of the variance of performance (Nash et al., 2000).

In the typical case when participants of virtual worlds are not on their own but rather interact with (and experience) others in the virtual, the feeling of being there is supplemented by the feeling of being with others (Schroeder, 2002; Schultze, 2011; K. M. Lee, 2004). Again, the fact that users are re-embodied and thus possess some sort of bodies seems to support the human brain in comprehending the actions of others (Bray & Konsynski, 2007). This points to the need of analyzing the mutual perceptions during social interaction (Bainbridge, 2007). Furthermore, because they incorporate social features that are necessary for these processes, virtual worlds are well suited to support knowledge creation and knowledge sharing; hence they are regarded as means for co-creation (J. Mueller et al., 2011; Kohler, Fueller, Matzler, & Stieger, 2011). Research findings indicate that, through social interactions, specific learning processes related to understanding the current state of a virtual world evolve in the context of its usage, which in turn further enhance social interaction and communication skills of users (Papargyris & Poulymenakou, 2005).

Some virtual worlds heavily rely on user-created content and action (cf. Cahalane et al., 2012). In contrast to other IT artifacts which are usually based on the work of few expert programmers, often very large numbers of users contribute to the content virtual worlds, thereby making up a large extent of it (Chaturvedi et al., 2011). An interesting example in this regard apart from Second Life is Minecraft. Its users avail themselves of blocks in order to build all kinds of structures. As the game developed of the years, users "worked together to create wonderful, imaginative things".[9] Also, skilled users sometimes try to modify virtual worlds in ways not originally intended by the manufacturer or developers, resulting in what is called modding (Schiele et al., 2011). Therefore a "symbiotic emergence of culture and content" of virtual worlds and their content has been detected which is unique to virtual world practices (Messinger et al., 2009, p. 205). Related therewith, the

9 https://minecraft.net/

rights of users over the content they create are being debated, however some see the higher risks on the manufacturer's side (Roquilly, 2011).

2.2.3 *Human Responses*

Users of virtual worlds are predominantly male (Barnett & Coulson, 2010; Ducheneaut et al., 2006), yet this may depend on the type of virtual world (e. g., Zhou et al., 2010; Jansz & Tanis, 2007) and be different with groups of higher age (Yee, 2006b). However, while avatars are on the move in the virtual world, the users who control them reside in the real world. As a consequence, real-world biases, prejudices, social norms, and biological mechanisms are likely to be transferred to and replicated in the virtual world (cf. Schultze, 2010; Ducheneaut et al., 2006), thus to regulate the behavior within, too (Cahalane et al., 2012). For instance, though they are free to choose from a number of different characters, users clearly prefer those that are pretty according to real-world criteria. Also, when male users choose roles with healing powers, they usually pick a female avatar. One can therefore assume that gender stereotypes are shared by the real and the fantasy world (Ducheneaut et al., 2006; Barnett & Coulson, 2010). Similarly, a team can achieve competitive advantage over another team by choosing to be represented by avatars with uniforms that are red rather than any other color; as it seems, the signal color red triggers similar reactions in human aggressive competition in a mediated situation as in real life and thus affects the outcome of competition in a similar fashion (Ilie, Ioan, Zagrean, & Moldovan, 2008).

Many technological issues of virtual worlds which relate to human factors are not fully solved yet. As for example, image quality seems to affect the presence experience and the level of immersion which a user can ultimately achieve (Bracken & Skalski, 2009). Furthermore, attention impacts the processing of biological motion in humans: When visibility is distracted while watching a surveillance tape, for example due to poor video quality, humans are liable to errors in interpreting biological motion correctly (Parasuraman et al., 2009). Presence is yet only enhanced "when body movements in interaction effects are not just arbitrarily coupled (a mouse-clickmoves the virtual body forward), but coupled in a way that fits the experiences one has with one's body" (Schubert, 2009, p. 170). A major challenge for avatar design is therefore related to computer graphics and animation and lies in finding an acceptable balance between complexity and control. As for instance, attempting to exactly control the position of objects in a virtual world and aiming at generating natural-looking movements at the same time are conflicting objectives (Sims, 1994). Additionally, latency effects may cause serious consistency problems (Bainbridge, 2007; Fritsch, Ritter, & Schiller, 2005). In this context, research on the appearance of and reactions to avatars or virtually embodied agents can draw on experiences with humanoid robots and an often cited paradigm called The Uncanny Valley (cf. Mori, MacDorman, & Kageki, 2012). The latter refers to the phenomenon that generally, the more human-like characters appear, the more they are liked–off to the point when the response to them completely switches from affection to discomfort. It is assumed that such a switch points to the fact that if a robot's appearance is very close to that of humans, even a slight mismatch between a robot's behavior and human expectations with regard to human-like behavior causes irritation (MacDorman, 2006b). Though not yet fully substantiated by empirical evidence (Gee et al., 2005), findings of several studies support this line of thought and highlight the need for a better understanding of behavioral realism rather than concentrating on achieving true visual realism (Groom et al., 2009; Tinwell, Grimshaw, Nabi, & Williams, 2011). As for instance, a recent study examined whether the uncanniness of an animated character, that is, the extent of how awkward a nearly human looking character is being perceived by others due to slight derivations from true human behavior–may depend on which emotion is being communicated by that character. The study showed that all emotion types are perceived as more uncanny when transmitted via nonhumans, yet that under certain conditions, the uncanny effect was stronger for some emotions (i. e., fear, sadness, disgust and surprise) than for other emotions (i. e., anger and happiness, cf. Tinwell et al., 2011).

Also concerning technological aspects, the question of how much sensory feedback a user can process has been raised (Stanney et al., 1998). Findings show that the rich experience offered by virtual worlds can also be a distraction, due to limitations in the information processing capacity and attention span of users (Irons et al., 2011; Nah et al., 2011). Furthermore, studies in the context of teaching indicate that rules of how and when to speak need to be established for communication to serve a particular purpose (Petrakou, 2010; cf. also Dennis, 1996). Also, virtual worlds are designed to produce a positive affect in users; however, the "fun factor" is not what makes an educational game successful in terms of learning performance (Kiili, 2005, p. 14). Interestingly, students that are liable to test anxiety have a greater sense of presence in virtual environments than in a neutral environment; moreover, high test anxiety students feel more presence than their non-test-anxiety counterparts (Alsina-Jurnet & Gutiérrez-Maldonado, 2010).

As with many technologies, practices, and applications in the virtual, trust and deception are major issues in the virtual worlds context (Xiao & Benbasat, 2011; Paul & McDaniel, Reuben R. , Jr., 2004; Pavlou & Gefen, 2004). Investigations in that field can yet lead to important insights (Korsgaard, Picot, Wigand, Welpe, & Assmann, 2010); especially the avatar-human interaction and the question of how users infer characteristics of the actual user behind an avatar (cf. Schroeder, 2002) are of significance in this regard. Related thereto, literature calls to further analyze humans' conceptualizations of their avatars or characters (Bainbridge, 2007). Concerning this, the appearance of an avatar seems to give valuable hints. As for instance, people express their own values by choosing avatars related to themselves and by decorating them (Suh et al., 2011). On the other hand, deceivers are more likely to choose avatars that are different from their real selves (Galanxhi & Nah, 2007).

An avatar's appearance has indeed shown to affect the cooperation between participants of virtual worlds in studies: Highest levels of cooperation were attained when avatars looked like a person rather than like a dog or a cartoon dog, even though the human-looking character and the characters resembling the dog and the cartoon dog had the same capabilities (Clayes & Anderson, 2007). Also, if avatars are introduced to virtual worlds users as being a user-controlled avatar rather than an AI-controlled agent (cf. also Schultze, 2010), users show greater physiological arousal to otherwise identical interactions during engagement in virtual worlds (S. Lim & Reeves, 2010). A study investigating the brain mechanisms that underlie the human perception of avatars found that on the surface, humans trust avatars in a similar way they trust other humans; on a deeper neurological level, however, there are notable differences with regard to cognitive processes related to mentalizing (Riedl et al., 2011). During mentalizing processes, individuals make inferences about the mental states of other agents in order and determine their future actions, often in an automatic manner (C. D. Frith & Frith, 2006). As these activities are not accessible to consciousness, researchers needed to inspect them with the aid of brain imaging (Riedl et al., 2011). Results of the aforementioned study were interpreted to the effect that the concept of a mind is not attributed to the same degree to an avatar as it is attributed to a human, but that nonetheless, "differences in brain activation do not lead to differences in trust behavior" (Riedl et al., 2011, p. 14).

It is not well understood what circumstances determine the avatar-user relationship and what makes users view their avatar as a "created object or thing, as an extension of self, or possibly as a child or offspring" (Wasko et al., 2011). Some clues are given by studies: People seem to react to avatars with enjoyment, and viewers' responses to them has been associated with art perception and aesthetics as well as social and emotional psychology (Klimmt, Hefner, & Vorderer, 2009). Users identification with their avatar increases the enjoyment of media products and the feeling of presence (Park et al., 2010). The user's sense of presence, in turn, increases with an avatar that looks like the user's actual self (Schultze, 2010). Also, the more the facial and bodily appearance of avatar and user are alike, the higher the levels of avatar identification, emotional attachment, and intention to use the avatar (Wasko et al., 2011).[10]

Little is known about how online behavior affects a users' behavior offline (Schroeder, 2002). Avatars seem to enable users to express themselves in ways they may not feel comfortable to express themselves in their real lives (Frostling-Henningsson, 2009). Achievements in eSports, for example, will not only be recompensed in-world, but typically lead to a gain in recognition and possibly financial profit in a user's real life (Schiele et al., 2011). Using virtual worlds are known to satisfy psychological needs for competence, autonomy, and relatedness (Przybylski et al., 2010), and leadership skills acquired in virtual worlds can be transferred to real-world scenarios (Barnett & Coulson, 2010). Also, virtual goods do not seem to fulfill any subsistence or physical need, but primarily satisfy a user's social need for self-presentation (Animesh et al., 2011; H.-W. Kim, Chan, & Kankanhalli, 2012). Users turn to virtual worlds in order to socialize (although male and female users "are looking for very different things in those relationships" Yee, 2006a, p. 774). Experiences of competence need satisfaction was found to be linked to greater immersion in the virtual world, increased self-esteem, and an increased likelihood of reengaging in the virtual (Przybylski et al., 2010). If engaged in virtual worlds over longer periods of time, users tend towards more stable virtual identities and respecting the social norms they have come to share with other members (Schroeder, 2002). Such consistent behavior creates the appropriate conditions to engender trust by proving the trustworthiness of the respective virtual world users (Mennecke et al., 2008).

In virtual worlds with many players which are connected at the same time, complex social interaction occurs, and rules and norms are needed to deal with conflicts. Often, such rules are partly developed by the publishers, but to a great extent, they develop among the participants themselves (Pargman & Erissson, 2005). Nevertheless, the duality of virtual world use as a "gate" between the real and the virtual worlds does not only reveal positive but also negative reciprocal effects (Yee, 2006c). The fact that users cannot actually see each other tends to foster disagreeable behavior in the virtual, causing little ethical concern on the part of the evildoer that is in control of the mediated action (Hundley & Shyles, 2010). Examples of such behavior include lying (Hundley & Shyles, 2010), deceiving (Galanxhi & Nah, 2007), and cheating (especially in context of eSports competition). Virtual world users are not afraid to "attack, swindle, thwart, and exploit each other" either (Barnett & Coulson, 2010, p. 167). Also, not all users are able to transfer encouraging experiences of self-efficacy obtained in the game into real life (Young, 2013), with the result that users may rather turn towards

10 This is a mechanism also known from sexual reproduction: Humans tend to prefer mates that resemble themselves–except in times of stress (Lass-Hennemann et al., 2010).

virtual worlds use than real-life activities to fulfill their psychological needs–possibly to escape from their real-life problems (Partala, 2011; Przybylski et al., 2010; Yee, 2006a). Modern psychotherapy recognizes that avatars are a key factor for addiction tendencies related to avatar-based gaming because avatars potentially allow users to compensate for–perceived–limitations in real life. Accordingly, a therapist needs to reveal what the specific character which the addicted virtual world user has created represents with respect to the user's personality and circumstances of life, for instance related to a perceived deficiency in self-efficacy in the actual living environment (Young, 2013; Barnett & Coulson, 2010).[11] Finally, just like end users of other information and communication technologies, virtual world are potentially exposed to technostress, cybersickness, and deleterious physiological aftereffects (Ragu-Nathan, Tarafdar, Ragu-Nathan, & Tu, 2008; Stanney et al., 1998).

2.3 PRIOR MEDIA THEORIES

As mentioned above, we are not interested in IS usage per se but rather in the outcome of IS usage and in performance in particular. As elaborated, virtual worlds qualify both as an IS and as a medium. In the following, we treat virtual world as the latter because the aspects of virtual worlds which are of interest to our study, namely avatar-user interaction and IR, relate to their media characteristics. Nevertheless it should be kept in mind that the IS-specific characteristics (e. g., the emergent nature of virtual worlds) are potentially important features related to performance as well, especially in work settings.

Two media theories are dominant in IS research, namely that of social presence and that of media richness (e. g., Dennis, Fuller, & Valacich, 2008). The next paragraphs shortly summarize these media theories and point out their deficiencies concerning their application to virtual worlds (cf. also Schiele et al., 2011). Hereinafter, we particularly refer to virtual worlds as IR media. However, virtual worlds are not the only IR media: avatar e-mail for example, numbers among them, too (cf. Y. Lee et al., 2009).

The presence of avatars enriches communication via IR media in a similar way as nonverbal cues enrich communication in general; as a result, IR media are considered as rich as video-conferencing (Y. Lee et al., 2009; Riedl et al., 2011). Avatar use also affects the experience of presence and supports communicator identification (Y. Lee et al., 2009). However, the ability of existing media theories to "explain the complex and dynamic interactions and events that unfold in real time within the persistent environments that are virtual worlds" has been

11 Studies indicate that dopamine levels of users are significantly higher during virtual world usage, a fact that may foster a physiological addiction; see http://www.video-game-addiction.org/what-makes-games-addictive.html.

called into question (Schultze & Orlikowski, 2010, p. 812). We briefly introduced social presence when talking about the concept of presence; as aforementioned, a medium's capability to provide social presence is adjudged by evaluating whether it is perceived as warm, personal, sensitive, and sociable (e. g., Animesh et al., 2011). The quality of social presence is therefore inherent to a medium. High levels of social presence are considered to be reflected in a high degree of salience of the person one interacts with as well as a high level of salience regarding the resulting interpersonal relationship (Yoo & Alavi, 2001; Y. Lee et al., 2009). In the context of virtual worlds, social presence has been defined as "the perception that there is personal, sociable, and sensitive human contact in the medium" (Saunders et al., 2011, p. 1085). Media richness, in turn, measures the richness of a medium according to four criteria: (a) its ability to give immediate feedback, (b) the variety of communication or social cues it allows to send with a message (like voice inflection, body gestures, words, numbers, and graphic symbols), (c) how features enable personalization of the medium (can personal feelings and emotions infuse the communication?), and (d) the language variety or range of meaning which is attainable (Daft, Lengel, & Trevino, 1987; Carlson & Davis, 1998).

The two theories are typically being consulted when effects of media or, related thereto, media selection are in the center of attention. The general assumption underlying these theories is that people choose media which best fit the task at hand, and that for tasks which involve ambiguous information, a richer medium is more appropriate, whereas less rich media should be applied to solve a task with a lack of information (e. g., Dennis et al., 2008). As the decision which medium fits a specific task (or communication objective) is assumed to be based on a rational choice, they are sometimes referred to as rational theories of media choice (Watson-Manheim & Bélanger, 2007; Y. Lee et al., 2009). Studies have linked social presence and media richness to various measures of performance as dependent variables–such as, for example, manager and team decision quality, decision time, message length, speed of response, consensus change, communication satisfaction, accuracy of information delivered, perceived equivocality, knowledge conversion, or sales (e. g., Daft et al., 1987; Dennis & Kinney, 1998; Sussman & Sproull, 1999; K. H. Lim & Benbasat, 2000; Miranda & Saunders, 2003; Massey & Montoya-Weiss, 2006; Brunelle & Lapierre, 2008; Y. Lee et al., 2009) as well as to observed group polarization processes (Sia et al., 2002). In turn, the theories have also been tested as the dependent variable, for example, in a setting where media conditions and group cohesion were hypothesized as antecedents to social presence or web site navigation and communication support (Yoo & Alavi, 2001; Zhu et al., 2010). Though they have developed independently, it has been observed that the theories of social presence and media richness are "surprisingly similar" (Carlson & Davis, 1998, p. 338; see also there for an overview of the historical development of the two research streams). As a consequence, they are often jointly applied to studies (e. g., Carlson & Davis, 1998; Yoo & Alavi, 2001; Y. Lee et al., 2009). In the context of IR media, Y. Lee et al. (2009) found that avatar e-mail provides higher levels of media richness and social presence than other types of media.

Both theories, however, have received substantial criticism. This criticism is mostly related to the fact that studies attempting to find empirical support for the theories' assumptions and implications reported mixed results, especially in the area of new media (Dennis & Kinney, 1998; Dennis et al., 2008). Additionally, media richness theory, for example, predicts differences in perceived media richness (Dennis & Kinney, 1998), a circumstance which stands in the way of classifying media in an incontrovertible fashion. As for instance, though e-mail was assumed to be a "lean" medium in terms of media richness, participants of a particular study used them for tasks that should require a richer medium. The researchers observed that, through certain practices, e-mails were simply redesignated, for example by sending socio-emotional content which made e-mails appear sociable and warm (Carlson & Davis, 1998). This points to the problem that the same technology may generally be received and used differently by different individuals, or that different receivers may vary in their skills to interpret cues and to communicate (Schiele et al., 2011). Social factors like group cohesion have already been found to add to media conditions and thus enhance social presence, irrespective of the medium used (Yoo & Alavi, 2001).

A study which explored determinants of the use of collaborative technologies for information sharing also found that the use of collaborative media is particularly influenced by task characteristics or task interdependence, respectively (Jarvenpaa & Staples, 2000). However, despite the fact that task types are often an important moderator of studies and that communication technologies have been found to reduce task conflicts (Wakefield, Leidner, & Garrison, 2008), the classification of task types is far from unambiguous or trivial (Niederman, Beise, & Beranek, 1996; Hwang, 1998; K. H. Lim & Benbasat, 2000; Dennis et al., 2008). The more recent approach of media synchronicity theory attempts to account for individual differences and social conditions by incorporating the familiarity of individuals with the tasks they are performing as well as with their coworkers, assuming that these factors affect the outcome of communication processes (Dennis et al., 2008). By postulating objective media capabilities, this theory is yet similarly liable to problems related to a media-dependent view. Furthermore, choices are always bound to the information available as well as people's limited capacities to process these information, or may be affected by individual goals (H. A. Simon, 1972; Rivard, 2014; K. H. Lim, Benbasat, & Ward, 2000).

Another important field which theorizes on media is communication research. This field has long concentrated on analyzing how mass media may influence opinions, attitudes, and actions, a research stream dubbed “the study of mass persuasion” (Katz, 1959, p. 1). More recent approaches have studied social and psychological change through, for example, incorporating psychological and social processes related to peer influence (Katz, 1959; cf. also Bandura, 2001). The question to be asked was therefore no longer “what do the media do to people?” but rather “what do people do with the media?” (Katz & Foulkes, 1962, p. 378). This represented a shift from a media-dependent perspective to a need-dependent perspective (cf. Yoo & Alavi, 2001). This shift gave rise to a new type of media theory, namely that of uses and gratifications. It represents a functional approach to media which contrasted previous “behavioristically oriented, stimulus-response type of theory”, but also included lines of thinking which were popular among “theorists of popular culture”–meaning humanists and psychoanalysts–back then (Katz & Foulkes, 1962, p. 379). Those who put forward this approach postulated that user of media consciously satisfy their needs through media use (Sherry, Lucas, Greenberg, & Lachlan, 2006). They also provided a classification of media-related needs in accordance with (a) three different modes which they linked to (b) four different types of resources a person can aim for, and looked at these relationships with respect to (c) seven possible referents including the self (Katz, Haas, & Gurevitch, 1973); this classification is displayed in Table 1.

Table 1: Classification of media-related needs. Adapted from Katz, E., Haas, H., & Gurevitch, M. (1973). On the Use of the Mass Media for Important Things. American Sociological Review, 38(2), p. 166.

	with respect to	
Mode	Connection	Referent
1. Strengthen	1. Information, knowledge, and understanding	1. Self
2. Weaken	2. Gratification, emotional experience	2. Family
3. Acquire	3. Credibility, confidence, stability, status	3. Friends
	4. Contact	4. State, society
		5. Tradition, culture
		6. World
		7. Others, negative reference group

In the course of their study, the authors further identified five groupings which, in their eyes, formed meaningful groupings. They are reproduced hereinafter:

- needs related to strengthening information, knowledge, and understanding–called cognitive needs;
- needs related to strengthening aesthetic, pleasurable and emotional experience-or affective needs;
- needs related to strengthening credibility, confidence, stability, and status–these combine both cognitive and affective elements and are labeled integrative needs;
- needs related to strengthening contact with family, friends, and the world; they can also be seen as performing an integrative function; and
- needs related to escape or tension-release which we define in terms of the weakening of contact with self and one's social roles.

As can be taken from our review of related work, the media-related needs just listed have not lost any of their relevance–even with new media like virtual worlds (Yee, 2006a). Related therewith, media theories incorporating social aspects are more topical than ever before (e. g., Ngwenyama & Lee, 1997; Miranda & Saunders, 2003; Hsu & Lin, 2008; A. Mayer, 2009; Duan et al., 2009; Zhu et al., 2010; Aral, Brynjolfsson, & Alstyne, 2012). Uses and gratifications has already been applied to study video game use and preferences (Sherry et al., 2006).

However, the framework has some conceptional shortcomings, making it not well-suited for the application to virtual worlds. First, it only refers to leisure settings and the achievement of personal goals rather than explaining what function media have in compulsory work settings. Second, it implies that individuals are aware of their needs and that they know about their need hierarchy, an assumption which is liable to issues of social desirability (cf. D. L. Phillips & Clancy, 1972). Third, uses and gratifications essentially considers a momentary snapshot of an individual's needs and their interrelations at a macro-level–without attempting to explain the outcome of media use. And fourth, though it has been observed that certain goals change with age (Katz et al., 1973), the approach does not account for the fact that media may create or affect needs or the need hierarchy, thus allow for interaction. It instead assumes that needs are competing and distinct attributes that serve particular needs, and that individuals need to chose media according to the needs they aim to satisfy. On a more general note, the approach is also inconsistent with its initial claim that it is contrasting the view on users as recipients of media solely: By assigning clear attributes to every medium and postulating the effect they have on an individual, the hypothesis that media have an distinctive effect on people–in that they receive, understand, and react to media similarly (Schiele et al., 2011)–is rather sustained than contradicted. Uses and gratifications has therefore also been called a subtradition of media effects research as opposed to a rigorous social science theory (Ruggiero, 2000).

In a nutshell, social presence, media richness, and uses and gratifications have mostly been applied to study media choices–thus the question that the theories actually address–rather than performance (cf. Dennis et al., 2008). Furthermore, all aforementioned theories (i. e., including media synchronicity theory) take, to a greater or lesser extent, a media-dependent perspective and thus suffer from conceptional shortcomings linked to that perspective: Differences in media effects, skills or experience, and social or contextual factors are disregarded (Yoo & Alavi, 2001; cf. also Animesh et al., 2011).

A media-dependent perspective is particularly unsuitable in the light of the large variety of virtual worlds that is available today (Bainbridge, 2007; Kohler et al., 2011; Schroeder & Axelsson, 2006). As noted before, similar arguments apply to the conceptualizations of tasks. To quote Carlson and Davis (1998),

> it is perfectly predictable that some groups or individuals will define either tasks or media traits differently, thus explaining the problems with media richness/social presence theories (pp. 339–340).

Furthermore, the perception of a medium's capabilities and appropriateness also depends on the circumstances of media use (Watson-Manheim & Bélanger, 2007). With regard to virtual worlds, we find that prior media theories particularly fail to deal with the distinctive characteristics of IR media, as we explain in the following. First, virtual worlds usually come with various communication channels that can be used simultaneously (Bartle, 2004). Voice and chat, for instance, are both an integrative feature of many virtual worlds. This provision of a combined use is not accounted for by traditional views on media, which typically assign voice and chat to separate media like telephone and instant messenger.

Second, thanks to re-embodiment and IR, virtual worlds transmit information that are unique to IR media, as they enable users to "express information that otherwise would not have been communicated" (Y. Lee et al., 2009, p. 451). Particularly, in addition to explicit, textual communication, virtual worlds provide means of expression which allow to transmit tacit knowledge, notably through the use of bodily representation (Schultze & Orlikowski, 2010; see also Nonaka & Takeuchi, 1995). In turn, virtual worlds' characteristics make them equivocal in the sense that they are accompanied by multiple, conflicting interpretations which require "hunches, discussion and social support" for individuals to deal with them (Berente et al., 2011, p. 686; citing Daft et al., 1987, p. 357). Virtual worlds hence allow to combine a large variety of mediated communication channels with elements of face-to-face meetings. This combined approach may ultimately help to encounter negative effects of exclusively verbal interactions on the one hand–namely that only small portions of information are being exchanged–and factors that inhibit information processing related to practices of group support systems usage–namely issues of information integration and credibility as well as of presenting important information in a conspicuous way–on the other hand (Dennis, 1996).

Third, the fact that all communication channels are conveniently combined in one medium not only facilitates their simultaneous use, it may also may overcome problems related to media supplying only single channels. It has been posited that members of organizations' communities share a common understanding of how and when certain media of combinations of media a appropriate to use in order to fulfill task like coordination, knowledge sharing, information gathering, relationship development, and conflict resolution; however, the fact that many different media exist and are being used side-by-side increases work complexity (Watson-Manheim & Bélanger, 2007). As all communication channels are integrated into one, the use of virtual worlds can reduce this complexity. Also, this feature fosters a customized usage of virtual worlds. Such use may thereby differ from that originally intended by the developers and producers of the virtual game; an example for this practice is modding (Schiele et al., 2011).

Forth, virtual worlds can be viewed as a form of multimedia. In the context of new media (such as multimedia), research has not been able to find convincing empirical evidence for the validity of, for example, media richness theory (Dennis, 1996), possibly because multimedia address many shortcomings of traditional media. As for instance, they were found to reduce equivocality of information related to ambiguous tasks, and multimedia representations are typically perceived as more useful than text-based representations (K. H. Lim & Benbasat, 2000). Also, humans are liable to so-called first impression bias, thus tend to make up their mind about something on the basis of the very first information that has been presented to them–a serious problem for organizational decision making. The use of multimedia can help diminish such effects (K. H. Lim et al., 2000). However, though multimedia are perceived as more challenging and engaging, they rather contribute to fun-aspects of an experience, and the technology itself draws a lot of a receiver's attention (Webster & Ho, 1997); this can be counterproductive with regard to learning success, for example. Referring thereto, virtual worlds represent a persistent, flexible, emerging, and socially interactive medium and storage for information, allowing users to take the time they need to process and think about information that is available to them, discuss it is with others, come back, and construct meaning in a collective effort (Dennis, 1996; Miranda & Saunders, 2003; Ngwenyama & Lee, 1997; Berente et al., 2011).

Fifth and most importantly concerning the objective of the present study, traditional perspectives on media do not address performance as such but rather explain media choices. At most they take an indirect approach to explain performance by looking, for example, at the outcome of media effects by matching a medium to task equivocality or perceived equivocality of information (Watson-Manheim & Bélanger, 2007; Dennis & Kinney, 1998; Dennis et al., 2008; K. H. Lim & Benbasat, 2000). Alternatively, classification of behavior related to new media (television, mobile phones, computers, game consoles and the Internet) is limited to variables such as frequency and variety of use, typical activity, and typical media platform, with a focus on satisfaction of media-related needs (Brandtzæg, 2010).

2.4 A THEORY OF INDIVIDUAL DIFFERENCES

The factors which frame the virtual world experience or foster the loyalty of virtual world customers are not only of interest to research (e. g., Animesh et al., 2011; Cao, Glukhova, Klamma, Renzel, & Spaniol, 2008; Choi & Kim, 2004; Ducheneaut, Yee, Nickell, & Moore, 2007; Verhagen, Feldberg, Hooff, & Meents, 2009; Wu et al., 2008; see also Bateman, Gray, & Butler, 2010; Moon & Sproull, 2008). As a matter of fact, the whole virtual world industry is constantly analyzing user responses to their products in order to improve virtual world design and customer loyalty–with great success. Flow and engaging experiences are essentially ”hardcoded” into virtual worlds, triggering criticism for intentionally designing virtual worlds so that users will become addicted to using them (Salen & Zimmerman, 2004, pp. 336–358).[12] Users of traditional means of communication have been found to become engrossed in communication through “telepresence, social presence, intrinsic motivations, playfulness, cognitive absorption, and flow” (Wasko et al., 2011, p. 648), yet new media like virtual worlds are particularly built to facilitate conversations, collaborations, and interactions in an engaging and fun way (q. v.). Consistent with this view, fun-aspects like enjoyment, flow, and cognitive absorption but also need satisfaction related to escapism, social interaction, and self-expression were found to be important concepts for virtual worlds use (Goel et al., 2011; Nah et al., 2011; Weiss & Schiele, 2013; Ducheneaut & Moore, 2004; Barnett & Coulson, 2010; Partala, 2011).

While we acknowledge that IT artifacts need to be used in order to improve performance (Goodhue & Thompson, 1995), we do not aim at further explaining virtual world use. Instead, we concentrate on the actual outcome of virtual world use.[13] In this regard, accounting for levels of presence, for example, has not contributed to explaining performance in the context of virtual worlds (Schultze, 2010), nor are we

12 See also remarks on “addictive properties” on http://en.wikipedia.org/wiki/Video_game_controversies.

13 It is noteworthy that recently, Nah et al. (2011) and Animesh et al. (2011) conducted two of the few studies that analyze outcome variables of virtual world use, namely brand equity and sales of virtual goods.

aware of any research associating enjoyment, flow, or similar constructs to individual human performance of virtual world use.

Related to a study-specific co-creation task of idea generation to be performed in virtual worlds, Kohler et al. (2011) stressed that, though the researchers were able to iteratively increase the level of participation and the amount of ideas contributed by users through design adaptations, "the outcomes for companies remain speculative and it remains unclear whether a co-creation system focused on the user experience outperforms other mechanisms such as monetary incentives to attract avatars" (p. 787). However, poor design is probably a significant constraint for performance:

> Human performance in VEs will likely be influenced by several factors, including task characteristics, user characteristics, design constraints imposed by human sensory and motor physiology, integration issues with multimodal interaction, and the potential need for new visual, auditory and haptic design metaphors uniquely suited to virtual environments. In order to maximize human performance in VEs, each of these factors must be considered (Stanney et al., 1998, p. 330).

So far, research on virtual worlds has tended towards analyzing them through qualitative rather than quantitative approaches (Bainbridge, 2007; see e. g., Hussain & Griffiths, 2009). Studies of virtual worlds have also concentrated on nontask-focused contexts, for instance when investigating the role of avatar application in social network services (Suh et al., 2011). The goal of the present study is to analyze individual performance as the dependent variable of interest. Our observations from literature as well as criticism related to prior theories can be summarized as follows:

- Arguably, differences in media choices are related to media capabilities and the task to be performed or the goals to be achieved. However, because both media capabilities and task characteristics are being judged differently by different individuals, media choices are always made from a personal perspective–or alternatively, from the perspective of a group of people who share the same social norms.

- By the same token, task characteristics are perceived differently by each individual. Additionally, users adjust their media usage (i. e., that of e-mail) to their current communication needs or task requirements. Assigning a task type or a particular need to a specific medium is therefore only suitable to a limited extent.

As expressed by Ngwenyama and Lee (1997), "it is through the process of enactment that people, not electronic communication media, bring about the richness that they experience in their communications"

(p. 146). Consequently, a conceptual approach which solely accounts for media types, needs, or task characteristics in order to predict performance will generate findings which are possibly not unambiguous or replicable, and even more so when dealing with the various types of virtual worlds that have emerged by now.

User characteristics and individual differences, on the other hand, are typically neglected by previous media theories, though individual cognitive and affective responses are already known to influence online experiences product purchasing, for example (Jiang & Benbasat, 2007). With regard to virtual worlds, this relates to the cognitive and emotional involvement with the medium (Schiele et al., 2011). Recent studies have confirmed the importance of cognitive and emotional aspects of the virtual world experience (Animesh et al., 2011; Wasko et al., 2011). Cognitive challenges related to navigation and knowledge acquisition during virtual world use have been found to affect task performance (Nash et al., 2000), pointing to the need for cognitive abilities to master virtual world tasks. Also, higher cognitive abilities should help identifying equivocality issues and finding adequate solutions to problems, or to deal with technical challenges such as design constraints. Furthermore, even experts from practice recognize that "emotional information" is being transmitted through avatars (Gartner, Inc., 2008). Navigation with others as well as voice chat enhances collaborative shoppers' perceptions of social presence during online shopping experiences (Zhu et al., 2010). Content appealing to emotions is also transmitted through the interaction with others, as virtual embodiment not only allows for constructing shared experiences, but also provides self-conscious observation and reflection (Schultze & Orlikowski, 2010).

Some phenomena, like individuals' levels of presence, immersion, and flow, thus defining sensations in the context of virtual worlds, are thought to be measurable and comparable, though this is still debatable (Sacau et al., 2008). Nonetheless, no coherent, unambiguous definition of the nature of the three concepts exists. Immersion is sometimes treated as a predictor, sometimes as a consequence of presence, and no empirical evidence for a causality of presence and performance was found so far (Schultze, 2010; Nash et al., 2000; Sacau et al., 2008). In turn, it has been suggested that various forms of (cognitive) absorption or (focused/psychological) immersion, which broadly refer to involvement, emotional engagement, or aspects of intrinsic motivation, potentially impact performance (Agarwal & Karahanna, 2000; Webster & Ho, 1997). Other virtual world phenomena, for instance humans' preference for avatars that resemble themselves, or mechanisms related to trusting avatars, are subconscious processes which defy self-reporting and need to be tapped into through other means (like affective startle response modulation or brain imaging, e. g., Park et al., 2010; Lass-Hennemann et al., 2010; Riedl et al., 2011).

The question of interest related to this study is: Given that virtual worlds are being used, which individual differences impact users' performance?

2.5 RESEARCH MODEL AND HYPOTHESIS DEVELOPMENT

Emotional and cognitive abilities as well as a user's experience are important in the context of virtual worlds, yet not all processes related to emotions and cognition which determine a virtual world experience are accessible to humans' consciousness, nor do they necessarily have an impact on IR performance. Since virtual world participants are controlled by users that reside in the real world, real-world theories which explain individual performance should hold when applied to virtual world users. Literature thereby points to the following:

- Dealing and aligning one's emotions in emotionally laborious interaction with others is an important ability which needs to be factored in when analyzing job performance (Joseph & Newman, 2010). Users embodiment and emotional involvement with avatars, and related to this, shared experiences with others in virtual proximity–often combined with the need to cooperate in order to perform–are likely to represent highly demanding conditions with regard to emotional capabilities.

- Cognitive ability is of great significance for individual performance. The cognitive, social, and situational requirements that virtual world tasks impose on users will challenge users' cognitive abilities at least as much as real-world tasks. This observation suggests that cognitive ability is an important factor for IR performance.

- Virtual world users are required to focus on the virtual representation of their own body and that of the other participants. Consequently, the less distracted they get through their real-life environment, the more they should be able concentrate on tasks (like anticipating others actions), thereby making less mistakes than in a less concentrated state (e. g., Rowe & McKenna, 2001; A. M. Williams, Ward, Knowles, & Smeeton, 2002). This points to a form of a deep involvement with the virtual world, also known as cognitive absorption (Agarwal & Karahanna, 2000).

- Users' prior knowledge and his or her experiences influence the way they experience and perceive the virtual world (Kiili, 2005). Through experience, technology becomes increasingly transparent (Schultze, 2010). The capability to understand the representations used, or rather to "read" and "write" in terms of IR media (Caperton, 2010), refers to a particular media literacy (Rogow, 2004).

These suggestions are further sustained by observations which indicate that in the context of virtual worlds, the most common user characteristics mentioned and analyzed are cognitive abilities, levels of perception and cognition, and experience level (Sacau et al., 2008; Stanney et al., 1998).

In IS research, emotions have been linked to IT use and adoption as well as to the perception of social presence during collaborative shopping experiences (Beaudry & Pinsonneault, 2010; W. W. Chin, Marcolin, & Newsted, 2003; Ortiz de Guinea & Markus, 2009; Zhu et al., 2010). Cognitive ability has been implemented in IS studies that examined training success (S. J. Simon, Grover, Teng, & Whitcomb, 1996), and, conceptualized as cognitive processing, mindshift learning (Armstrong & Hardgrave, 2007). A common view is that the actual processing of data, "at least in the arena of managerial communication (...)[,] is performed not by the hardware or software, but by the human beings themselves" (Ngwenyama & Lee, 1997, p. 146). The role of individual capabilities related to media literacy or experience with a medium with regard to individual performance has largely been neglected in IS research.

However, certain emotional capabilities and cognitive ability are well-studied in various disciplines and have been linked to many performance or success measures. The impact of cognitive ability in multiple facets on job performance or academic achievement is also a classic field of action in psychology and organizational research (e. g., J. E. Hunter & Hunter, 1984; J. E. Hunter, 1986; LePine & Dyne, 2001; Lyons, Hoffman, & Michel, 2009; McGrew & Wendling, 2010; Newton & McGrew, 2010), and literature constantly produces adjusted measurements according to newest developments in research (Loevinger, 1957; McGrew, 2009; Flanagan & Harrison, 2012). Emotions have been associated with general work performance (Elfenbein & Ambady, 2002; Côté & Miners, 2006; T.-Y. Kim, Cable, Kim, & Wang, 2009; Kuncel, Hezlett, & Ones, 2004), group performance (Ashkanasy & Daus, 2005; J. R. Kelly & Barsade, 2001), decision making (Mikels et al., 2010), effectiveness of negotiation (Elfenbein, Foo, White, Tan, & Aik, 2007), attitude towards organizational change (Vakola, Tsaousis, & Nikolaou, 2004), consumer responses (Bagozzi, Gopinath, & Nyer, 1999), prosocial behavior (Bagozzi & Moore, 1994), drop-out rates in higher education (Kingston, 2008), and bodyweight regulation (Perugini & Bagozzi, 2001), to name a few. Some studies have also investigated the importance of emotion alignment for tasks that require different (verbal vs. spatial) cognitive abilities (Storbeck, 2012). Conversely, cognitive ability is a significant predictor for processing emotional stimuli and emotional information (Fiori & Antonakis, 2012). In this regard, cognitive absorption is known to enhance focusing on events which are being mediated, similar to an experience of flow and related to the feeling of control (Agarwal & Karahanna, 2000). Being cognitively absorbed should thus enable a user of virtual worlds to concentrate more on events, the environment, other users and their emotions, as well as their own emotions, and should reduce distraction from the task at hand (Schiele et al., 2011).

Media literacy is relatively new movement which mainly emerged with the diffusion of new media (G. Schwarz, 2005). Generally, media literacy acknowledges the unique experience of new media interactions (Burke, 2008). In an attempt to unify different views of various streams, the (U.S.) National Leadership Conference on Media Literacy (Aufderheide & Firestone, 1993) defined a media literate person as someone who can "decode, evaluate, analyze and produce both print and electronic media" (p. 1). They have also created media literacy standards which suggest which competencies should be taught to children (Burke, 2008). Related abilities already begin to form during childhood, a process which sets the course for later media understanding (Yan, 2006; Christakis & Zimmerman, 2009; D. R. Anderson & Hanson, 2010). Such skills refer to the ability to access, analyze, evaluate, and communicate messages in a wide variety of forms (A. M. Rubin, 1998; Hobbs, 1998).

Overall, previous research seems to indicate that a model which attempts to explain IR performance should account for individual differences that relate to emotional capabilities–possibly intensified through the effects of cognitive absorption (Schiele et al., 2011), cognitive ability, and a special kind of IR media literacy. Our research model (depicted in Figure 2) borrows from psychology, developmental studies, media literacy as well as IS and incorporates the four previously mentioned concepts as predictors of IR performance.

In the following, we present the domain conceptualizations of our constructs and our hypotheses development. First, we outline the conceptualizations of the primary constructs, followed by those of the secondary constructs (for more information on this distinction see Section 1.3 and Section 2.1), partitioned according to the line of thinking they are assigned to. Following MacKenzie et al. (2011), aspects to consider when developing a conceptualization of a construct are (a) examining on how the focal construct has been used in prior research or by practitioners, (b) specifying the nature of the construct's conceptual domain, (c) specifying the conceptual theme of the construct, and (d) defining the construct in unambiguous terms. These specifications are required to "exacting in delineating what is included in the definition and what is excluded" (Churchill, 1979, p. 67), aiming at construct clarity by (a) parsimonious distinctions between concepts, (b) delineating scope conditions, (c) showing semantic relationships to other related constructs, and (d) demonstrating logical consistency of the construct with the theoretical argument (Suddaby, 2011, p. 347). Some constructs are distinctively contextualized with a focus on the IR concept, in order to signalize the relevance of activities or experiences that share aspects of both the real world and that of a virtual world. A

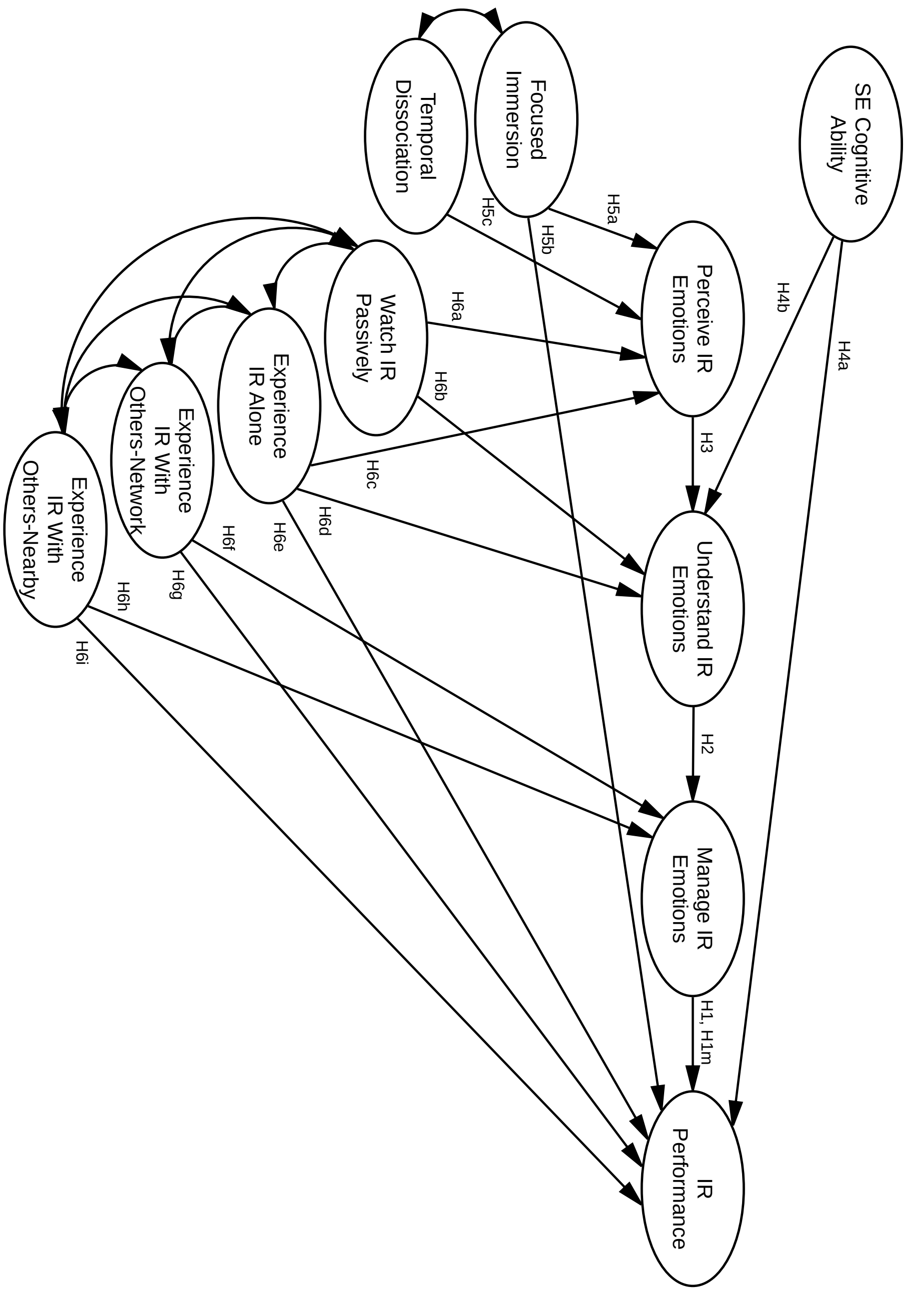


Figure 2: Hypothesized structural model of IR performance.

summary of the conceptualizations and hypothesized dimensionality of all constructs is provided in Table 2 at the end of this section (p. 74).

In order to decide whether to opt for higher-order factor structures or to generally avoid them, we developed an iterative approach: First, we reviewed the possible advantages of their implementation. We therefore examined the guidance of when to decide in favor of a second-order model supplied by Wetzels et al. (2009), F. F. Chen, Sousa, and West (2005, pp. 473–474), or in the "checklist" by Hair et al. (2009[14], as cited by Weiber & Mühlhaus, 2010, p. 224). A position taken by many authors is that a researcher needs to establish clearly and in a well-argued fashion why a higher-order conceptualization is more suitable than conceptualizing correlated factors (e. g., Bagozzi & Yi, 2012; Weiber & Mühlhaus, 2010). Generally, it is assumed that a higher-order construct should represent an highly abstract concept, with the lower-order factors representing its less abstract, specific components that refer to a common event or target, or which influence each other (Bagozzi & Yi, 2012). As for instance, hierarchical models of intelligence are well established, whereas assuming a general factor for personality facets is not considered appropriate (Bühner, 2011, p. 420). With some of the constructs included in our investigations, the discussion regarding this issue has already taken place, and literature provides according and well established suggestions (as an illustrative example, see the paragraph on cognitive absorption). As a next step, we reflected upon our constructs once again, in order to detect compelling reasons for a higher-order conceptualization of any of them. We concluded that assuming correlated factors was adequate for all our constructs. However, for good measure and cross-validation, we decided to additionally take a higher-order structure into account where previous conceptualization or statistical evidence would be inconclusive.

2.5.1 *Emotions, Cognitive Ability, Cognitive Absorption, and IR Performance*

The interrelationship between emotions, cognitive ability, and individual performance has been studied extensively in the past; however, it is quite complex (Joseph & Newman, 2010; Côté & Miners, 2006; Ashkanasy & Daus, 2005; Bastian, Burns, & Nettelbeck, 2005).

To begin with, origin, nature, and purpose of emotions have always been subject to controversy (Dewey, 1895). Even among today's top researchers of scientific psychology there is disagreement on the concept of emotion, on formal criteria to determine what qualifies an emotion, or on what causes or succeeds an emotion (Russell, 2003). Some have suggested that an emotion refers to an integrated, short term feeling

14 Hair, J.F./Anderson, R.E./Tatham, R.L./Black, W.C. (2009): Multiple Data Analysis, 7th Edition, New Jersey (as from the references of Weiber & Mühlhaus, 2010, p. 299.

state “including happiness, anger, or fear, that mix varying amounts of pleasantness-unpleasantness and arousal-calm, among other sensations” (J. D. Mayer & Salovey, 1997, p. 23). From the point of view of Joseph and Newman (2010), the term emotion applies to “surprise, joy, anger, sadness, as well as state affects of moods”, the latter covering pleasant and unpleasant state affect (pp. 55–56), while according to Beaudry and Pinsonneault (2010), the list of emotions studied in IS includes excitement, happiness, anger, and anxiety.

Early studies only accounted for two dimensions of emotions, referred to as valence and arousal. However, some have recently argued that the components of emotions account for many more facets, namely (a) appraisals of events, (b) psychophysiological changes, (c) motor expressions, (d) action tendencies, (e) subjective experiences, and (f) emotion regulation (Fontaine, Scherer, Roesch, & Ellsworth, 2007). As some emotion-related processes are considered nonconscious and automatic (spontaneous facial expressions, motor-preparedness, and physiological changes) and some are expected to be linked to conscious appraisal of the self or a situation, particular focus lies on trying to determine what effects of emotions are controllable through the conscious mind (Schubert, 2009; J. D. Mayer, Roberts, & Barsade, 2008; see also Kapoor, Burleson, & Picard, 2007). The ultimate goal of such considerations is to examine whether (or how) humans can deliberately control or use their emotions and whether such capabilities can help achieving a goal or increasing performance (e. g., Elfenbein & Ambady, 2002; Elfenbein et al., 2007; Côté & Miners, 2006; Fiori & Antonakis, 2012; Storbeck, 2012).

In 1990, Salovey and Mayer published their initial and much-cited concept of what they termed emotional intelligence (EI), a framework which integrated (a) the appraisal and expression of emotion in oneself and others, (b) the regulation of emotion in self and others, and (c) the use of feelings into a new construct. EI thereby involves the ability to carry out accurate reasoning about emotions and the ability to use emotions and emotional knowledge to enhance thought (J. D. Mayer et al., 2008, p. 507). The revised definition of EI by J. D. Mayer and Salovey (1997) involved (a) perception, appraisal, and expression of emotion, (b) emotional facilitation of thinking, (c) understanding and analyzing emotions as well as employing emotional knowledge, and (d) reflective regulation of emotions to promote emotional and intellectual growth (p. 11).

Since then, the EI concept has received a lot of attention in research and nonscientific literature (Joseph & Newman, 2010). Many of the various more recent conceptualizations typically rather cover the four capabilities of emotion perception, emotion regulation, emotion understanding, and emotion utilization instead (Ciarrochi, Chan, & Caputi, 2000). The “diversity in the conceptions of EI” (J. D. Mayer et al., 2008, p. 509) that have been employed has been identified as one of the main causes for the continuing disagreement regarding the topic amongst researchers (Ashkanasy & Daus, 2005; cf. e. g., Schutte et al., 1998; and E. J. Austin, Saklofske, Huang, & McKenney, 2004, for a criticism), though several integrative approaches have been developed (see J. D. Mayer et al., 2008, for a discussion). The debate mainly concentrates on whether or not EI qualifies as an elusive construct or intelligence, respectively, on its own (Law, Wong, & Song, 2004; Brackett & Mayer, 2003; J. D. Mayer, Caruso, & Salovey, 1999), and if so, whether it is constituted by distinct subdimensions (e. g., J. D. Mayer, Caruso, & Salovey, 2000), or whether it is to be interpreted as one intelligence dimension among others and viewed as linked to general intelligence (cf. Becker, 2003; MacCann, 2010).

While part of the criticism of EI is based on the observation that it shows conceptual redundancy with cognitive ability (Johnson, Johnson, & Roseth, 2012), Locke (2005) criticized the general concept of an emotional "intelligence". He argued that, in his opinion, (a) the ability to monitor one's emotions did not require much intelligence, that (b) discriminating between emotions was a learned skill, and that (c) the everyday use of one's knowledge was not an issue of intelligence as such. Nonetheless, he acknowledged "the importance of one element of EI in human life: introspection" (p. 429), which he summarized as the ability to "identifying the contents and processes of one's own mind".

Furthermore, broader conceptualizations, referred to as mixed-model approaches (Joseph & Newman, 2010; J. D. Mayer et al., 2008), showed overlapping traits with personality when tested against each other (Law et al., 2004; Davies, Stankov, & Roberts, 1998; Brackett & Mayer, 2003). This has raised two issues, namely (a) the scope of the concept (broad vs. narrow) and whether (b) the competencies in question represent a *trait* (e. g., Tett, Fox, & Wang, 2005; Cooper & Petrides, 2010; Freudenthaler, Neubauer, Gabler, Scherl, & Rindermann, 2008) or an *ability* (cf. Daus & Ashkanasy, 2005; J. D. Mayer et al., 2000; Iliescu, Ilie, Ispas, & Ion, 2013). As a result, the distinction between the two constructs of trait EI and ability EI was made. Petrides and Furnham (2003) described the former as emotion-related self-perceptions and dispositions and the latter as objective emotion-related ability, and though their theoretical domains overlapped, the concepts differed fundamentally in their operational definition (i. e., self-report questionnaires vs. maximum-performance tests); eventually, the authors came to the conclusion that "self-reports must be given priority over objective measures in the study of affect" (p. 52), favoring a conceptualization as a trait characteristic.

The construct of objective cognitive ability or, respectively, "psychometric" intelligence (Furnham, Moutafi, & Chamorro-Premuzic, 2005), is also the subject of highly controversial debate (Freund & Kasten, 2012). The conceptual theme of the construct revolves around the capacity of an individual to process information (see, e. g., Colquitt, LePine, & Noe, 2000). Many–overlapping or competing–theories exist about the possible structure of cognitive ability and its different manifestations (Beier & Ackerman, 2005), and researchers argue to which extent it is an innate, stable individual difference or rather dependent on age (Benson, Hulac, & Kranzler, 2010), gender, experience, or education within a certain culture (Furnham et al., 2005). The historical development of general intelligence assessment and the structure of ability constructs since the 1890s is summarized by Ackerman and Heggestad (1997). Widely adopted references treating theories, tests, and issues of ability assessment aim at catching up with recent advances and pushing the development process forward with every new edition (e. g., Flanagan & Harrison, 2012).

Our extensive examination of the PsycARTICLES[15] and PsycINFO[16] databases revealed that many authors who incorporated cognitive ability in their study did not explicitly clarify (if not neglect) what was included and excluded in their definition of the construct nor why. However, the use of a specific measurement instrument inherently *implies* a clearly outlined definition, for it is determined by the underlying theory of the instrument. However, by reading the description of the measures used, one can assume which ability dimensions the authors intended to measure. For example, Côté and Miners (2006) explained that they requested their candidates to find similarities or differences in figures; Furnham and Chamorro-Premuzic (2004) reported that their measures included "word and number comparisons, disarranged sentences, serial analysis of geometric figures and story problems that require mathematical and logical solutions" (pp. 153–154); B. S. Bell and Kozlowski (2002) assessed individuals' "mathematical, verbal, logical reasoning, and spatial ability to create a measure of general mental ability" (p. 501); and S. J. Simon et al. (1996) tested a subject's aptitude via questions that targeted the following areas: "(1) ability to understand instructions, (2) potential for learning a job quickly, (3) ability to solve problems, and (4) ability to come up with new ideas and new work directions" (pp. 477–478, numbering in original).

To "quick fix" conceptualization issues, authors highlight that the chosen–often only commercially available–instrument has been used "widely" so far. Nonetheless, these instruments may not be in alignment with latest theoretical advances. One of such a widely used instrument is the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). However, Benson et al. (2010) found that a Cattell-Horn-Carroll (CHC) structure described the underlying constructs of the instrument much better than the structure proposed and published in its technical and interpretive manual by its own developers. At present, the CHC theory is referred to as "the most comprehensive and empirically supported psychometric theory pertaining to the structure of human cognitive

15 www.apa.org/psycarticles
16 www.apa.org/psycinfo

abilities" (Benson et al., 2010, p. 122). In its current form, it proposes a model of nine broad and more than 70 narrow abilities (Newton & McGrew, 2010; see also McGrew, 2009, for a comprehensive overview).

In response to a lack of orientation, researchers yet began to make use of many ability tests in one single study, sometimes using a whole battery of up to 12 tests and more (see, e. g., Bastian et al., 2005; or Ackerman & Wolman, 2007). This has lead to the emergence of the CHC-based cross-battery assessment approach (XBA), in an attempt to provide a more systematic and a theoretically and psychometrically sound approach to "crossing" assessment batteries (Flanagan, Alfonso, & Ortiz, 2012).

Interestingly, while the latest developments allow for an assessment of "general" intelligence exhibited through a broad range of cognitive abilities, they also allow researchers to focus on instruments which only assess a narrow cognitive construct of interest to the particular study (Flanagan et al., 2012). We therefore concentrated on abilities that have been considered important for the field of IS before. Related therewith, IS research has associated general cognitive ability with problem solving skills (Highsmith, 1978), the capacity to learn (Webster & Martocchio, 1992), or the ability to perform a mindshift (e. g., from one programming paradigm to another, see Armstrong & Hardgrave, 2007).

Many theoretical are based on the assumption that humans differ in terms of cognitive and attentional abilities, information processing capacities, and capabilities pertaining to emotions. Furthermore, they assume that an individual's pool of these resources is limited and that allocation strategies to cope with requirements of the environment may affect achievement in terms of performance (Joseph & Newman, 2010; Dennis, 1996; Irons et al., 2011; Nah et al., 2011; M. Y. Yi & Davis, 2003). Broad definitions of performance, however, are rare. In scientific literature, performance, if at all, is usually not conceptualized to its full extent but rather narrowly operationalized in terms of the applied performance measure; one of the few counterexamples in this context is job performance, which has been defined as the degree to which an individual has contributed to an organizations goals (Côté & Miners, 2006). In turn, a large variety of indicators of performance have been applied in different studies so far, as for example:

- the level of perceived equivocality (K. H. Lim & Benbasat, 2000);
- task-oriented and socioemotional outcomes of meetings and meeting results in an organizational context (i. e., meeting outcomes like directly measurable results of a meeting, members' overall satisfaction or satisfaction with particular aspects of meetings as well as long-term impacts of a meeting, see Niederman et al., 1996);
- task-performance as effectiveness (in terms of displaying knowledge or skills, communication, taking charge, etc.) and organiza-

tional citizenship behavior directed at the organization as well as at individuals (Côté & Miners, 2006);

- points scored in a negotiation exercise (Elfenbein et al., 2007); and finally,

- skilled anticipation during sequences of action in a tennis game (Rowe & McKenna, 2001).

In the context of performance in virtual environments, it has been argued that measures need go beyond task outcome due to their complex nature (Stanney et al., 1998; see also Schultze, 2010).

With regard to predictors of performance, findings are mixed. While general cognitive ability, sometimes referred to as *g*, has been found to be the best single predictor of job and learning performance–especially with difficult and complex tasks (J. M. Phillips & Gully, 1997; B. S. Bell & Kozlowski, 2002), specific knowledge like experience has notable value in this regard, too (Ree, Earles, & Teachout, 1994). Also, cognitive ability seems to be unsuitable to predict sport performance of professional National Football League (NFL) athletes or computer training success in nonvoluntary settings such as a U.S. Navy training. Lyons et al. (2009) concluded that cognitive ability was not a substitute for strength, speed, endurance, and agility, thus qualities which are needed to fulfill the physically challenging tasks performed in the NFL. S. J. Simon et al. (1996) suggested that hands-on experimentation with computers and practical application of knowledge may require different skills than those needed for general comprehension tasks which involve thinking and less procedural knowledge.

Recently, Joseph and Newman (2010) performed a meta-analysis in order to test a cascading model of emotional abilities which adapted theoretical considerations of the EI literature and also incorporated cognitive ability. They found empirical evidence for a causal chain of emotion perception, emotion understanding, and emotion regulation preceding job performance. Also, their findings substantiated the importance of cognitive ability for emotion understanding as well as for job performance. Emotion regulation had more impact for jobs which required more emotional labor, such as jobs with a high amount of contact with clients and customers (Daus & Ashkanasy, 2005). The results of the moderator analysis are depicted in Figure 3. As elaborated above, we integrated and adapted those parts of the model which covered aspects with relevance to our study, namely emotional and cognitive abilities. We built our model of IR performance on this basis, thereby drawing on research hypotheses which have been empirically validated.

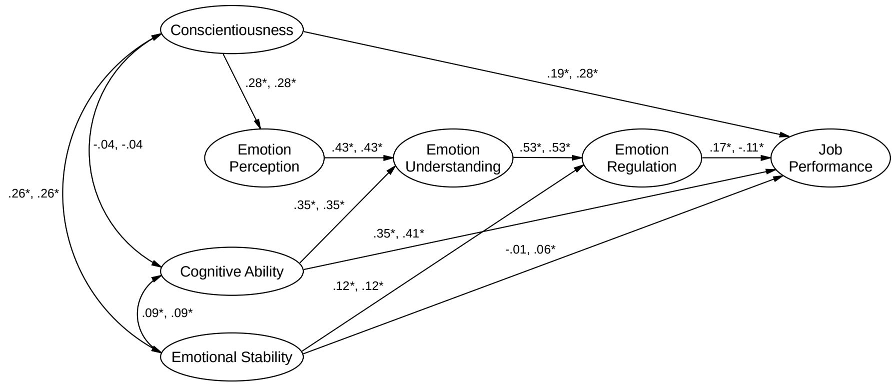


Figure 3: Results of the post hoc analysis by Joseph and Newman (2010) on the basis of their cascading model of emotional intelligence (EI), which the latter tested through meta-analysis on data from performance-based EI measures. Emotional labor (EL) was postulated as a moderator of EI; parameters of high EL are presented first, followed by low EL parameters. Adapted from: Joseph, D. L., & Newman, D. A. (2010). Emotional Intelligence: An Integrative Meta-Analysis and Cascading Model. Journal of Applied Psychology, 95(1), p. 70.

As emotional stability exhibited very low and insignificant effects related to cognitive ability and job performance throughout the whole study, we disregarded this construct. Also, as our study concentrated on the effect of emotions and because the emotion subfacets involved here only refer to conscious emotion mechanisms, we also excluded conscientiousness, an antecedent to emotion perception and job performance in Joseph and Newman (2010)'s model, in favor of a less complex model.

However, several circumstances made the implementation of objective instruments to measure cognitive ability unsuitable for our study: (a) Even in their shortest versions, a certain time is needed to answer all questions of such a test, (b) no control over cheating can be enforced unless under laboratory-like conditions, and last but not least, (c) acceptance of this test among our survey candidates was expected to be very low. Therefore, we abandoned the implementation of objective cognitive ability originally planed and decided to integrate self-estimated (SE) cognitive ability into our model instead. In this regard it is noteworthy that the term "construct" is often carefully avoided when referring to self-estimated or self-perceived ability measures, as opposed to "variable" or "concept". Freund and Kasten (2012) "globally define[d] self-estimation as a person's perception of her or his own abilities" (p. 299) and further described it as a *process* of repeated assessment experiences which eventually leads to domain-specific ability self-estimates. The subject is closely related to the construct of self-concept, that is, the theory that an individual has constructed about himself (Epstein, 1973), because self-estimation is assumed to be an *expression* of an individual's self-concept (Freund & Kasten, 2012). It has been argued that in this light, the exact delineation of what constitutes intelligence played a subordinate role when it is self-assessed, yet that instead, it mattered how a person evaluates his or her own abilities with regard to other people (Freund & Kasten, 2012).

The history of assessing subjective ability is similarly long as that of objective ability. The examination of the relationship between self-estimated, self-other, and objective measures of cognitive ability goes back to the early 20th century (Ackerman & Wolman, 2007). To grasp a person's self-estimation of ability, researchers have often drawn on asking their participants to compare themselves to *others*, requesting them, for example, (a) to rate their own abilities when compared with other people in the population at large (Ackerman & Wolman, 2007) or (b) to indicate how they thought their performance had been in comparison to the performance of their average classmate (or neighbor, friend etc., cf. Ames & Kammrath, 2004; Dunning, Johnson, Ehrlinger, & Kruger, 2003). With regard to these approaches, one can make an important distinction (Ackerman & Wolman, 2007): While the first approach aims at self-estimated intelligence as a more general concept, the latter captures self-evaluated performance in an exam–right after an actual test experience (Furnham et al., 2005). The distinction is important because (a) the cultural background has an impact on people's understanding of what being intelligent means, thus people from different cultures will have different things in mind when comparing themselves to an average person of the population and (b) test performance does not necessarily reflect the "true score" of psychometric ability, as exam nerves and other factors may bias the performance (Freund & Kasten, 2012). Another aspect worth considering is that when being asked to compare oneself to others, the outcome may be biased by the so-called "better-than-average" (BTA) effect (Krueger & Mueller, 2002); the BTA effect describes the phenomenon that people believe that they are better and that they do better than the average person. But as the approach of using self-estimates has shown reasonably accurate results (Ackerman & Wolman, 2007; Greven, Harlaar, Kovas, Chamorro-Premuzic, & Plomin, 2009), we adopted this strategy for our study. To incorporate a certain correction of the BTA bias, our approach was to let survey candidates self-estimate their abilities through the eyes of others, asking them how *others* had estimated the candidate's individual achievement in different areas in the past.

The third construct on which our model builds, cognitive absorption, can somewhat be interpreted as an intensifier of an individual's capability to perceive emotions in IR as well as an enhancer of IR performance. As introduced to the IS literature by Agarwal and Karahanna in 2000, this construct describes an individual difference of a person that uses an IS. The authors circumscribed cognitive absorption essentially as a situational intrinsic motivator and as a state of deep involvement with technology–a characteristic which changes over time, rather than being stable. The construct derives its theoretical bases from the personality trait dimension of absorption (Tellegen & Atkinson, 1974), cognitive engagement (Webster & Ho, 1997), and Csikszentmihalyi's state of flow (1990), the latter being a construct from positive psychology with relevance to the subject of optimal performance (see, e. g., Fullagar & Kelloway, 2009; Jackson et al., 2001). Evidence of Agarwal and Karahanna (2000) seemed to indicate that cognitive absorption is an antecedent to perceived ease-of-use of an IT, and that in turn, playfulness and personal innovativeness, both individual traits, have strong effects on cognitive absorption (Agarwal & Karahanna, 2000). Some of the findings regarding the antecedents to cognitive absorption were later contrasted by Goel et al. (2011), who found empirical evidence that in the context of virtual worlds, rather three specific states of awareness, that is, social, location, and task awareness were predictors of cognitive absorption.

As noted by Agarwal and Karahanna (2000), their approach provides an additional perspective to a strictly utilitarian one and attempts to account for the fact that experiences with IT become more and more pleasurable and enjoyable trough richer and more appealing technologies, thus making IT usage actually intrinsically motivating (Agarwal & Karahanna, 2000). Users of digital devices often report a sense of temporal displacement during usage (Hundley & Shyles, 2010).

However, one of the construct's general limitation is that the state of cognitive absorption is measured retrospectively and is not examined during the actual activity or immediately after it (Agarwal & Karahanna, 2000). An earlier and similar construct is the state of playfulness, sometimes equated with the flow state, which integrates elements of the subjective experience associated with the engagement with an IT artifact; it "incorporates the extent to which: (a) the user perceives a sense of control over the interaction with the technology, (b) the user perceives that his or her attention is focused on the interaction, (c) the user's curiosity is aroused during the interaction, and (d) the user finds the interaction intrinsically interesting (Webster & Ho, 1997).

IR PERFORMANCE Performance in research literature often refers to "the level to which a person has learned to perform a particular skill" (J. D. Mayer & Salovey, 1997, p. 22). Moreover, performance is frequently circumscribed as the achievements or outcome of an individual in a particular area. Although the construct is regularly conceptualized in a broad sense, actual conceptualizations of performance are usually narrowly defined, and operational definitions are often closely tied to the context of a study. Côté and Miners (2006), for example, assessed job performance of individuals through their supervisor's rating of these individuals' task performance, subdivided into categories like effectiveness in displaying knowledge and communication skills. According to other authors, job performance covers task performance and contextual performance, where task performance refers to "an assessment of individual task output in terms of its effectiveness, i.e., the degree to which it meets the task goals" and points to behaviors required to fulfill a task, whereas contextual performance points to job-independent behaviors which help to set the social and psychological conditions require to perform a job (Burton-Jones & Straub, 2006, p. 235). Frequently, performance is simply conceptualized through a single achievement indicator. Examples of this practice are numerous: To grasp performance, researchers have relied on results of specific computer tasks (S. J. Simon et al., 1996; B. S. Bell & Kozlowski, 2002), ratings of the uniqueness of a proposed solution to a problem (Garfield, Taylor, Dennis, & Satzinger, 2001), results of memory tests after a memorization procedure (Bowman et al., 2009), the number of correct answers to a question as well as answer speed (Y. Lee et al., 2009), kickboxing performance scores at tournaments (Devonport, 2006), and last but not least, game performance in a custom-made video game (Bösche, 2009).

In media research, performance has been attributed to task performance, interpersonal communication, and equivocality (e. g., Dennis et al., 2008), attainments which are essential in IR because virtual worlds are subject to the social interaction and action orientation of their users (Schiele et al., 2011). Accordingly, we searched for a suitable concept to reflect IR performance of individuals from our target group (i. e., the community of the Internet portal we examined) and identified the user's overall gaming performance as appropriate for our purposes. We therefore define IR performance as the achievements of an individual in IR which is reflected in the so-called player level. In line with examples from literature, we considered the construct to be unidimensional and decided to collect objective information for this construct rather than self-reported performance information.

EMOTIONS AND IR PERFORMANCE IT artifacts are known to trigger emotional reactions from individuals in many ways (Beaudry & Pinsonneault, 2010). With regard to virtual worlds in particular, it has been suggested that imaginal and emotional responses of users are capable of explaining their acceptance of this technology (Holsapple & Wu, 2007). Positive emotions have been associated with telepresence, enjoyment, brand equity, and behavioral intention in the context of virtual worlds (Nah et al., 2011). However, though social influence is believed to be an important factor for virtual world usage, results may depend on the genre of the virtual world investigated, as gaining status, the size of the network, and normative beliefs like interpersonal influence–drivers which have already been identified in this context–may not foster usage of all virtual worlds to the same effect. However, benefits of social interaction and seeking for human contact and feeling of belonging seem to have a relatively stable influence (Mantymaki & Riemer, 2011).

Contents of consciousness can appeal to emotions irrespective of how this content has come into consciousness. Because of the power of human imagination, the same effects on core affects can be shown for films, plays, novels, and music as can be shown for real-life events, such that "it is not difficult to create situations in the laboratory in which people confuse mental imagery with reality" (Russell, 2003, p. 155). Knowledge from emotion studies is thus applicable to virtual worlds, as IR experiences are likely to take place on consciousness levels which influence core affects. Gaming virtual worlds which tell background stories of the characters to appear in the game receive a higher evaluation due to emotional, motivational, and physiological responses of their users (Park et al., 2010), and avatar identification positively affects the emotional attachment to one's avatar (Suh et al., 2011). People who observe a–physical or mediated–body read the emotional states, intentions, and personality traits by an empathic simulation of them (Biocca, 1997). Due to the constant need of dealing with others, users of virtual worlds are often required to engage in emotional labor–and are particularly required to react in a controlled fashion, in order not to get distracted from the actual performance goal. It is assumed that emotional labor in real life is highest in jobs with a high amount of contact with clients and customers, such as in customer service occupations and the helping professions (Daus & Ashkanasy, 2005). Individuals differ in their capability of selecting an emotion regulation strategy that prevents them from being distracted and helps them managing the personal resources they have at hand in order to maintain overall job performance (Joseph & Newman, 2010).

In literature the concept of EI typically relates to "a set of interrelated abilities possessed by individuals to deal with emotions" (Wong & Law, 2002, p. 244). As to the question of how many dimensions of EI can be ascertained, Brackett and Mayer (2003) found support for four, whereas others have stated that evidence for a four-factor structure was lacking (Becker, 2003). Davies et al. (1998)'s critique of EI measures even concluded that, of the many factors associated with the construct, the only candidate really entitled to be considered distinct was emotion perception. However, J. D. Mayer et al. (2000) argued that laboratory experiments showed that perception of emotion in the self correlates significantly with the ability to assess emotions in others. In a more recent publication, Joseph and Newman (2010) tested a three-dimensional sequential–or cascading–model of EI (as depicted in Figure 3) which excluded the fourth dimension hypothesized by J. D. Mayer and Salovey (1997), emotion facilitation, "due to its increasingly well-known conceptual redundancy with other EI dimensions and its lack of empirical support" (p. 55). The model thus incorporated the three subfacets of what the authors termed "performance-based" EI: emotion perception, emotion understanding, and emotion regulation.

With regard to recent developments and the purpose of our study, we more generally refer to emotional *capabilities* in an IR setting. Emotional capabilities thereby cover the ability of a person to perceive and to deal with his or her emotions (Law et al., 2004) or with those of others (cf. Elfenbein & Ambady, 2002) at a conscious level (J. D. Mayer & Salovey, 1997; Joseph & Newman, 2010). Following Tett et al. (2005), who posit that the two approaches trait EI and ability EI are complementary rather than contradictory or even mutually exclusive, we do not conceive these skills as purely ability-based, but expect them to include some aspects of disposition instead. We thus leave the discussion revolving around the term "intelligence" aside and rather look at a continuum of competencies which are, at one end, "stable, dispositional capacities that are either genetically endowed or acquired over a long period of socialization" and, on the other end, "narrow, highly specific capacities to achieve high-level performance with respect to a particular activity or task .... acquired (...) through observation or training" (Scherer, 2008, p. 103).

Many studies which have implemented the abovementioned (or very similar) emotion facets with regard to outcomes like job performance point to the conclusion that the use of emotions has predictive power (Law et al., 2004). They found that if a leader can accurately appraise the current emotional situation of their followers and are able to influence their followers' emotions, he or she will more easily reach them and make them receptive for their vision or an organization's goals (Pillai, Williams, Lowe, & Jung, 2003). The role that emotions play in terms of performance can be explained through how understanding and regulating of one's emotions as well as understanding of others' emotions affect working with others, since these abilities "are the core factors affecting intrapersonal well-being and interpersonal relations" (Law et al., 2004, p. 486). For example, anticipating social interruptions has been associated with reduced stress and increased task performance (Carton & Aiello, 2009). One finding of Joseph and Newman (2010) was that jobs with high demands in terms of emotional labor benefited from certain emotional capabilities, namely the ability to perceive, understand, and regulate emotions. They also suggested that emotional capabilities "positively predict (...) performance for high emotional labor jobs and negatively predict (...) performance for low emotional labor jobs" (Joseph & Newman, 2010, p. 54), meaning that individual emotional capabilities seem to matter even if they are not necessarily needed for a particular task.

In line with findings discussed above, we conceptualize IR emotional capabilities as a multidimensional domain (cf. Tett et al., 2005) and ground them on the cascading model supported by findings of the meta-analysis by Joseph and Newman (2010, see above). In order to reflect a sequencing or causal chain of the three, the latter suggested the three facets emotion perception, understanding, and regulation to be standalone constructs. Accordingly, our model accounts for the capability to (a) perceive one's own and others' emotions, (b) understand one's own and others' emotions, as well as to regulate one's own emotions; however, we slightly alter or adapt the regulation facet by additionally including the ability to act on others' emotions, thus capturing the ability to (c) manage emotions in oneself and others (cf. Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006) rather than just regulate one's own emotions.

Emotion regulation refers to the "*conscious* [emphasis added] regulation of emotions to enhance emotional and intellectual growth" and has further been defined as "the processes by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions" (Joseph & Newman, 2010, p. 55), thus pointing to self-regulation. As aforementioned, we assume virtual world users to engage in high levels of emotional labor due to the requirements imposed by the presence of various other people sharing the environment–either through direct social interaction or other social interruptions that may occur at any time of the task performing period. We therefore expect emotion regulation to play an important part in order to explain individual IR performance. In turn, acting on others' emotions can be defined as the ability to influence the current emotional state (or mood) of a person using one's own perception of that person's emotions and understanding how to influence them. It represents the capability of an individual to regulate one's own emotions for the purpose making that person change his or her current emotional state to a (desired) different state. This capability can be used to either interfere with the other person's performance (negative influence) or to engage in cooperation with the person (positive influence). Both strategies are important in virtual worlds and both lead to increasing one's own performance, depending on the current situation. We thus assume this capability to influence IR performance, too. A concept which in fact covers these two aspects of emotion regulation is emotion management. According to Brackett et al. (2006), "managing emotion pertains to the ability to reduce, enhance, or modify an emotional response in oneself and others, as well as the ability to experience a range of emotions while also making decisions about the appropriateness or usefulness of the emotion in a given situation" (p. 781). Due to the fact that virtual worlds require to constantly deal with other participants to reach a certain goal in terms of performance, we consider including the capability of managing emotions in oneself and others (rather than just the ability to regulate one's own emotions) to be a promising extension of Joseph and Newman (2010)'s model in the context of interactive virtual worlds. Inside of virtual worlds, emotions are mediated through reembodied users and users themselves reside in the real world (cf. Vincent & Fortunati, 2009; Schwartz, 2011). For an IR setting, however, there are two possible application areas for managing one's emotions, either not directly noticed by the other remote participants of the virtual world but mediated due to disembodiment, or alternatively, in interaction or in competition with co-users or contestants in close physical proximity. Thus both aspects of managing emotions should impact performance.

H1: Managing IR emotions is positively related to IR performance.

The general theme of a virtual world may affect the importance of emotional capabilities with respect to performance. Joseph and Newman (2010) suggest that the effect of emotions on job performance is moderated by the level of emotional labor involved in a job. Different levels of emotional labor may analogously be observed in different virtual worlds. This implicates that the effects of emotion management should then differ across different groups.

H1m: Game genre moderates the effect of IR emotion management with regard to IR performance.

Emotion understanding refers to the cognitive appraisal of emotions. In analogy to general cognitive ability, it can be interpreted as an individual's knowledge base and skills at any given point in time related to the origin and the consequences of emotions, thus to the understanding of how certain emotions transition from one another and how emotional experiences affect individuals (Joseph & Newman, 2010; Brackett et al., 2006). As such, it "entails understanding how emotions evolve over time, how emotions differ from each other, and which emotion is most appropriate for a given context" (Joseph & Newman, 2010, p. 57) and influences how an individual responds to emotion. Following the findings of Joseph and Newman (2010), it precedes emotion regulation, and by the same token, it should also precede acting on others' emotion; consequently, emotion understanding should precede emotion management, because it provides the rational for any subsequent action of a person. However, different to their initial hypothesis (Hypothesis 2, p. 58), Joseph and Newman (2010) did not find emotion understanding to fully mediate the effect of emotion perception on emotion regulation (though the direct effect was smaller than the indirect effect, see p. 65). We hypothesize that this causality and type of effect also holds in the context of virtual worlds and IR; users of virtual worlds are not only be required to understand mediated emotions, but also those emotions expressed by a fellow co-user in close physical proximity, who happens to use the same virtual world and with whom they cooperate or compete (e. g.,during an on-site eSports competition).

H2: Understanding IR emotions is positively related to managing IR emotions.

Emotion perception as treated by Joseph and Newman (2010) refers to the ability to perceive emotion in the self and others as part of the same construct, namely emotion perception, and captures the ability to identify emotions not limited to in oneself and other humans, but also in other stimuli "including voices, stories, music, and works of art" (Brackett et al., 2006, p. 781). When it comes to understanding emotions, individuals who aware of the verbal and nonverbal cues of their surrounding and their own emotional state have an information advantage over those who are unaware (Joseph & Newman, 2010). The capability of an individual to accurately perceive another person's emotions, that is, the type and intensity, facilitates the prediction and understanding of that person's subsequent actions (Brackett et al., 2006). We postulate that the causality of events, namely the perception of emotions preceding emotion understanding, is the same for IR. Additionally, since virtual world users sometimes share the shame space in the real world (like the same room), users may not only be required to perceive and be aware of emotions which are being mediated, but also transmitted through real-world cues by co-users or contestants.

H3: Perceiving IR emotions is positively related to understanding IR emotions.

It was to be revealed through analysis whether a progressive structure in which one emotional capability precedes another was valid, and whether support for the facets hypothesized was to be found; as expounded above, the topic is still controversial in literature.

SE COGNITIVE ABILITY AND IR PERFORMANCE This study focuses on abilities important to the IS context. However, as explained above, conducting an objective ability test was not suitable. We therefore needed to consider a valid approximation, which we found in subjective ability assessment.[17] Achievement feedback relevant to IS covers feedback on astuteness, capability of comprehending and learning quickly, and general capacity, which we therefore included. In order to also account for the fact that cognitive ability is connected to experience and education (e. g., Furnham et al., 2005), we additionally included cultural and institutional facets, namely the user's educational background, intellectual giftedness, and academic achievements.

A large body of literature supports the significance of general cognitive ability for achievement at work as much as for achievement at school (Kuncel et al., 2004). For several reasons, we believe that performance in the context of tasks which are being supported by IT like virtual worlds profits from cognitive ability. On the one hand, virtual worlds are a highly innovative yet equivocal technology, which challenges its users to make sense of the limited information they provide (Berente et al., 2011). Concerning the aspect of innovativeness, it has been suggested that IT innovators and adopters are equally important

17 Burson, Larrick, and Klayman (2006) found that judges, independent of their skill levels, are subject to similar degrees of error.

to implement them, namely to generate novel ideas yet to incorporate them into existing work structures as well (Garfield et al., 2001). However, related to sense-making and the construction of meaning, virtual worlds are inherently performative (Schultze, 2011). Findings from IS research suggests that cognitive ability is a good predictor of general comprehension tasks which require more abstract thinking and less procedural knowledge compared with hands-on experimentation (S. J. Simon et al., 1996). Once an individual has acquired procedural knowledge about how to perform a cognitive activity, this individual will develop knowledge and refine it in order to organize and access this knowledge; the outcome of a learning process which involves actively engaging in symbolic coding, cognitive rehearsal, and reproducing the newly acquired skills is likely to lie in higher levels of declarative knowledge and self-efficacy (cf. M. Y. Yi & Davis, 2003). Factors which are also assumed to contribute to performance in virtual worlds relate to spatial and strategic capabilities because of their importance regarding the navigational complexity of the virtual environment and related task (cf. e. g., Stanney et al., 1998). Some have expressed the view that individuals with less developed spatial capabilities may have difficulties to deal with the mental representation of the mediated environment (Sacau et al., 2008).

Due to its broad importance for performance and due to the nature of tasks in virtual worlds which require complex, abstract knowledge, we thus propose that SE cognitive ability–as reflected by achievement feedback from others (during daily experiences, school, etc.)–to have a positive impact on IR performance.

H4a: SE cognitive ability is positively related to IR performance.

On the other hand, emotion literature points to the necessity of intellectual or cognitive capabilities in order to deal with the requirements imposed by social interactions. As for instance, group discussions–which are comparable to interactions with others in virtual worlds–engage their participants in a large variety of information-related processes like information recall, information exchange, and information processing and use, all a the same time (Dennis, 1996). The influence of cognitive ability on emotion understanding is also indicated by several studies from psychology (Joseph & Newman, 2010). It is the interpretation of a situation or event, thus the result of conscious appraisal and rationale, which induces emotions in humans like anger or fear (Epstein, 1973). We can therefore assume that an individual which has received positive feedback related to his or her cognitive ability is likely to be capable of processing more emotional cues and to better understand emotions in him- or herself and others from the real and the virtual environment than an individual with has received less such feedback.

H4b: SE cognitive ability is positively related to understanding IR emotions.

Though we accounted for several facets in order to capture the achievement feedback users have received during their life, they were nonetheless not considered distinct factors of a higher-order construct nor standalone dimensions.

COGNITIVE ABSORPTION, EMOTIONS, AND IR PERFORMANCE

The conceptualization of cognitive absorption by (Agarwal & Karahanna, 2000) is problematic because it mixes the personality trait curiosity (cf. Jennett et al., 2008) with situational sensations like temporal dissociation measured "retrospectively" (Agarwal & Karahanna, 2000, p. 688). This is further sustained when studying the theoretical bases of the construct: According to Agarwal and Karahanna's definition, cognitive absorption is a second-order construct, exhibited through the five dimensions (a) temporal dissociation, (b) focused immersion, (c) heightened enjoyment, (d) control, and (e) curiosity. Due to restrictions of the analytical tool they used, the authors were unable to model the indicators for these dimensions, but claimed to have found sufficient support for the existence of five distinct factors within their results (Agarwal & Karahanna, 2000, p. 688).[18] However, in contrast to the assumption that all factors are dimensions of the same construct, Wakefield and Whitten (2006) argued that the dimensions of control, curiosity, immersion and dissociation represented cognitive functions, while enjoyment should rather be considered an affective construct; hence they measured enjoyment separately for their study.

Similarly, though they based their conceptualization of cognitive absorption on Agarwal and Karahanna (2000) and cited the latter to have modeled "cognitive absorption as a reflective higher-order construct, consisting of four dimensions" (p. 758),[19] for their study of virtual worlds, Goel et al. (2011) only took the dimension of temporal dissociation into account, thereby deferring from a higher-order model. They justified this approach by arguing that (a) their conception of cognitive absorption focused on temporal dissociation and the user's experience, and that (b) measuring any one of the dimensions was basically equivalent to measuring the other, as the cognitive absorption's reflective lower-order constructs correlated with one another. Concerning (a), the authors relied on previous studies in the context of video games, which associated time loss experience with immersion and positive experience with technology (Wood, Griffiths, & Parke, 2007). Concerning (b), the authors followed Burton-Jones and Straub (2006), who measured cognitive absorption along the single dimension of focused immersion; the latter argued similarly that the five dimensions were interchangeable

18 Note that the authors applied PLS. They stated that using CBSEM would "permit a closer examination of the dimensions of CA as second order factors" (Agarwal & Karahanna, 2000, p. 688).

19 In fact, the model actually consisted of five dimensions (see above); a precursor consisting of three dimensions was also presented by Agarwal, Sambamurthy, and Stair (1997).

and that in this case, aiming at simplicity was possibly even more valuable than aiming at "completeness of conceptualization" (p. 237).

We too tend towards the views expressed by Goel et al. (2011) and Burton-Jones and Straub (2006) on the subject. For our conceptualization, we thus first evaluated which aspects of cognitive absorption are considered essential in the context of our study–that is, which aspects best capture the extent to which a user is absorbed when using a virtual worlds system–and to then concentrate on them only. To this end, we centered on temporal dissociation, as it has been found to be important in the context of virtual worlds in particular (Goel et al., 2011), and on focused immersion, as it accounts for the impression of being absorbed (Burton-Jones & Straub, 2006).

Cognitive absorption is generally considered an intrinsic motivator and captures a user's employment of a system user and how a user is absorbed when using it (Wakefield & Whitten, 2006; Burton-Jones & Straub, 2006). The construct relates to psychological absorption, a personality trait which is defined in terms of openness to experience emotional and cognitive alterations, being prepared to respond to new stimuli, and to try new activities in a variety of situations (Wakefield & Whitten, 2006). With regard to our study, absorption "permits the user to easily suspend disbelief about the new virtual world presented by the media and, thus, forget the real one" (Sacau et al., 2008, p. 2263); moreover, it has been associated with spatial presence (q. v.). Descriptions of cognitive absorption further incorporate the capability of total cognitive or attentional engagement and flow (Agarwal & Karahanna, 2000; Wakefield & Whitten, 2006), the latter of which is known to be strongly entangled with telepresence (Nah et al., 2011) and an important predictor of game playing (Hsu & Lu, 2004; for a discussion of flow as a state or trait construct, see Marsh & Jackson, 1999). Noteworthy in this regard is the observation that presence as such has not been found to contribute to performance (Schultze, 2010). However, it can be assumed that the degree of presence may somewhat be linked to cognitive absorption by enabling a smooth, uninterrupted flow or progress of work.

One of the two facets of cognitive absorption we included in our examination, namely focused immersion, has been associated with the extent to which a user exploits the features of a system in order to perform a task, and has specifically been found to positively affect performance in cognitively engaging tasks which can profit from IT application (Burton-Jones & Straub, 2006; Agarwal & Karahanna, 2000). We assume tasks in virtual worlds to match this description and focused immersion thus to have a positive influence on IR performance.

H5a: Focused immersion is positively related to IR performance.

Cognitive theory related thereto assumes that cognitive functions generally precede affective responses (Wakefield & Whitten, 2006). We thus assume that the cognitive functions of cognitive absorption which we included, namely focused immersion and temporal dissociation, both precede the perception of IR emotions. With regard to focused immersion, the type of concentration which comes with this aspect of absorption should allow a user to become deeply involved with the virtual world, its inhabitants and their actions, and in particular with their emotions as well as one's own emotions.

H5b: Focused immersion is positively related to perceiving IR emotions.

In contrast to the effect-relationship postulated by Wakefield and Whitten (2006), Goel et al. (2011) developed and tested a model which conceptualized cognitive absorption as a *consequence* of three distinct types of awareness rather than as their *antecedent*. The three awarenesses, namely social awareness, location awareness, and task awareness, were modeled such that they refer to "users' awareness of whom they interact with and how they interact within a virtual world, what they interact about, and where, in a virtual sense, such interaction occurs" (p. 749). The authors argued that cognitive absorption describes a mental state resulting from "what users encounter, what they observe, what they perceive, and what their mental state is" (p. 750). Their further suggested that their results may serve as a basis for the reconceptualization of the construct of cognitive absorption. Though the authors first measured all five dimensions of cognitive absorption, thereby applying the measures adapted from Agarwal and Karahanna (2000), they finally only included the facet of temporal dissociation into their analysis, proposing to capture "a state of deep involvement that a user experiences as she performs an activity in the V[irtual]W[orld] and tends to lose track of time" (p. 752).

We generally agree that temporal dissociation can also be conceptualized as a result of an IT-related activity (this argumentation also indirectly stresses another possible limitation of the application of cognitive absorption in its original conceptualization). Furthermore, we suggest that temporal dissociation has neglectable value as a predictor of performance. However, Goel et al. (2011) decided to measure temporal dissociation, thus "the inability to register the passage of time while engaged in interaction" (Agarwal & Karahanna, 2000, p. 673), as a measure for a mental state which is reached through awareness. Though temporal dissociation may incorporate aspects of need satisfaction or escapism, an assignment of this dimension to mentalizing–as suggested by the interpretation of temporal dissociation as a consequence of an activity–is not compelling and has not been expounded sufficiently. While we support the idea of an involvement which engages the interest and the emotion of a user, we believe that any type of dissociation, thus the "temporary disruption in consciousness, memory, identity or perception of the environment" (Sacau et al., 2008, p. 2266), is not well suited to reflect a mental–thus somewhat conscious–state which results from "perception and awareness of virtual artifacts" as suggested by Goel et al. (2011). We thus interpret temporal dissociation as a standalone facet of cognitive absorption. In our view, it does not necessarily reflect a performance-oriented dimension, yet can occur in parallel while being cognitively absorbed. It should therefore highly correlate with focused immersion and support the perception processes in virtual worlds through blanking out the need to keep track of time in the real world.

H5C: Temporal dissociation is positively related to perceiving IR emotions.

In favor of a parsimonious model (cf. also Wong, Law, & Huang, 2008), we left considerations concerning antecedents and determinants of cognitive absorption aside. Also, as we only accounted for two factors of cognitive absorption, a higher-order solution was not eligible (Weiber & Mühlhaus, 2010, p. 220; Kline, 2011, p. 249). Similar to Wakefield and Whitten (2006) and Agarwal and Karahanna (2000), we therefore conceptualized the two factors as intercorrelated factors of one construct. We had to await results obtained along the study to find out whether what was believed to represent different dimensions would turn out to capture slightly different facets of the unidimensional cognitive absorption construct, or whether we would find substantial support for different distinct yet correlated factors of a multidimensional construct; with regard to the discussion in literature summarized above, both factor solutions seemed plausible.

2.5.2 *Childhood Experience and IR Media Literacy*

The " ability to create personal meaning from verbal and visual symbols we take in every day through television, radio, computers, newspapers and magazines, and of course, advertising" (Thoman, 1999, p. 50) is an aspect of media literacy. Many scholars and educators agree that media literacy, though relatively new, is an essential field of research today (G. Schwarz, 2005). Concerning the interest of the present study, we believe that especially the equivocality of virtual worlds and the type of messages transmitted via re-embodiment and mediated social interaction calls for a specific type of literacy closely related to IR (Berente et al., 2011; Schiele et al., 2011). A definition of media literacy with practical relevance in this regard has been given by A. M. Rubin (1998): "Media literacy, then, is about understanding the sources and technologies of communication, the codes that are used, the messages that are produced, and the selection, interpretation, and impact of those messages" (p. 3). It is generally assumed that experiences with the same media vary across different individuals such that the sense that one makes out of mediated messages differs; children can yet learn certain skills that help them in this regard (Thoman, 1999). Consequently, childhood experience forms the basis of later IR media literacy.

It is known that children at the age of five to nine rarely watch TV on their own but rather in the presence of siblings most of the time (Haefner & Wartella, 1987). This coviewing is assumed to enhance television's effects (Gentile, Nathanson, Rasmussen, Reimer, & Walsh, 2012). Similarly, studies on youth video game play showed that 76% of young video gamers play with others at least some of the time (Bers, 2010); and finally, computer games played via Internet, "particular virtual multiplayer games, come to satisfy the human need for community and social interaction" (Bers, 2010, p. 148). Further studies confirm that older children use media and digital devices to improve their social skills, satisfy their social needs or to even manage their social life (Hundley & Shyles, 2010). The early years of childhood and adolescence thus have an imprinting effect on later media capabilities and the understanding of the social and emotional processes related to media (D. R. Anderson & Hanson, 2010; Bers, 2010). Furthermore, experiments show that children undergo particular developmental stages important for their spatial capabilities at the age around seven. They also begin to make a distinction between objects and social agents, as they become aware of the latter's spatial perspective in a setting or a scene; for example, they make use of that social agent's perspective on an object to describe an object's position (Surtees, Noordzij, & Apperly, 2012). This underpins the importance of social aspects not only for the development of emotional, but also for spatial competencies in the media context, especially with regard to shared virtual worlds. Users of virtual worlds which are not familiar with this type of medium are assumed to be less capable on focusing their attention in the content of the media world; they may also succeed less well in assigning their attention to more than one stimulus (Sacau et al., 2008).

Despite the significance of media experience for the development of essential competencies, D. R. Anderson and Hanson (2010) nonetheless found that in media consumption studies, for example, on television habits, grasping the extent of children's experiences with media is often reduced to a simple "concept of 'watching television' " (p. 240), that is, television's effects are only measured in terms of hours a child sits in front of the TV screen. Other aspects, for example, mediation by another person or other facets of a child's TV experience and knowledge (on image, audio, comprehension of content, transitions from one shot to another, etc.), are usually not accounted for.

In literature, researchers investigating childhood media experience most often deal with actual children, and they usually interrogate them or their parents, respectively, to evaluate their media consumption.[20]

20 Note that parents usually underestimate their children's amount of media consumption, for example, in the case of television (cf. e. g., Cho & Cheon, 2005). As a result, parents' estimates may not reflect the actual experience correctly.

By contrast, our study targeted grown-up subjects, so that we needed a different approach than previous studies. Ultimately, we requested our participants to remember their IR media habits and experiences (video games, consoles, etc.) as a child at the age of 10 to 14 years. This particular age interval was chosen for several reasons: On the one hand, understanding of the technical complexity predates understanding of the social complexity of the Internet at a younger age (Yan, 2006); also, with the age of 14, childhood "ends" and typically, a new stage of life begins. On the other hand, our pilot studies showed that participants had difficulties remembering things that had happened before the age of 10. Overall, our strategy seemed the most suitable in order to estimate a participant's childhood inter-reality media experience (CIRME). Hence, similar to the approach of measuring cognitive ability via self-reported measures, we asked our participants to compare themselves with friends or relatives in their childhood to be able to rate the amount they consumed IR media back then.

For our purposes, we conceptualize the construct of CIRME as the degree to which an individual encountered IR media during childhood and the extent to which it was a social experience. The construct thereby builds on developmental psychology and media literacy as explained above. It captures the experiences and the understanding an individual has acquired while learning how to handle virtual worlds and to deal with related social interactions with others as a child. To account for the fact that coviewing and coplaying seem to be as important for the development of certain emotional skills as the "consumption" of the medium when experiencing an IR situation, we modeled four dimensions to cover the different types of experiences that can occur when using IR media: IR media use in the form of (a) passively watching others involved in IR actions, (b) experiencing IR on one's own and alone, (c) experiencing IR with others via network (e. g., connected through the Internet), and (d) experiencing IR with others in physical proximity (e. g., in the same room).

Observing others' interactions puts a child in a position to learn which emotional cues expressed by other children or adolescents–friends as well as siblings and their friends, but also through the mediated characters on the screen–lead to what type of emotional reactions–without being emotionally involved oneself. A child is thus being enabled to develop an effective strategy regarding the own emotional resources in such a situation, and can interpret it without own emotions interfering with the learning process (cf. Storbeck, 2012). We thus assume that this observation position not only positively affects the capability to perceive, but also to understand IR emotions.

H6a: Watching IR action passively is positively related to perceiving IR emotions.

H6b: Watching IR action passively is positively related to understanding IR emotions.

The exploration of IR media without anyone watching and interfering allows children to experiment with the medium and one's own emotional reactions to it, without having to deal with others' emotions and responses or behavior which may disturb the learning process. Though the social component is missing, this condition should significantly contribute to the perception and understanding of one's own IR-related emotions. The active role the user can play under such circumstances, thus actually participating in the virtual world and its tasks, will also positively affect that user's IR performance.

H6c: Experiencing IR alone positively related to perceiving IR emotions.

H6d: Experiencing IR alone positively related to understanding IR emotions.

H6e: Experiencing IR alone positively related to IR performance.

Though emotional responses of others are being mediated, when experiencing others in the virtual world, certain capabilities with regard to perceiving and understanding IR emotions are still needed to master the situation. Owing to the fact that other users of the virtual world are remote and nobody is present to observe, the individual can now give a trial to managing and particularly acting on others' IR emotions, without necessarily fearing to fail regarding the regulation of his or her own emotions. However, emotional responses may become visible through mediation, thus the capability of managing IR emotions of an individual is likely to increase thanks to experiences made under this condition. The participation in virtual world tasks, thus sharpening the user's mind with regard to the medium, also strengthens the performance of the individual in IR.

H6f: Experiencing IR with others via network is positively related to managing IR emotions.

H6g: Experiencing IR with others via network is positively related to IR performance.

The most challenging situation with regard to IR emotional capabilities is a one where not only other participants cooperate or comped with oneself in the virtual, but also in the real world, for example, if some or even all users are gathered in the same room. This experience mainly requires not only to regulate one's own IR emotions but also to act on others' IR emotions, whether they are being mediated or experienced in a real-life context. Active participation in virtual world tasks thereby positively affects the user's IR performance.

H6h: Experiencing IR with others in physical proximity is positively related to managing IR emotions.

H6i: Experiencing IR with others in physical proximity is positively related to IR performance.

At this stage, theoretical arguments seemed to speak in favor of a–more parsimonious–first-order model with four correlated factors, yet we considered to test for a higher-order factor structure along the analysis as well.

2.5.3 *Other Effects and Differences Across Groups*

This section presents constructs and related hypotheses we have considered potentially important in previous publications (cf. Schiele et al., 2011); they are, however, not part of the research model of the present study. We have previously referred to them as our secondary constructs and secondary hypotheses, respectively. The following paragraphs elaborate on their significance for explaining IR performance and develop our secondary hypotheses. They also cover some additional considerations regarding differences across groups which could possibly exist as well as control variables which have the potential to be of interest to the present study. Due to restrictions in time and space, they were not tested as part of this thesis.

SELF-MOTIVATIONAL TRAITS The setting of higher goals has been found to lead to increased performance across a variety of tasks (J. M. Phillips & Gully, 1997). By the same token, achievement motivation is believed to amplify the positive effects of goal orientation on actual performance (see Yperen & Janssen, 2002, also for an overview of the different labels for this concept). If the perceived consequences of performing a particular action are rewarding, this will encourage an individual to repeatedly perform well in a task (M. Y. Yi & Davis, 2003). Hence self-regulation is not only a concept related to EI; in fact, the ability of self-regulation (in conjunction with the possession of self-reflective and self-reactive capabilities) is also important in order to exercise control over thoughts, feelings, motivation, and actions–through anticipation of future events and according behavioral standards (Bandura, 1991). It has been proposed that so-called procrastinators, thus people who tend to delay the accomplishment of all tasks, put certain effort in managing their emotional reactions to the situation and protect their self-esteem by giving themselves an external reason for failing; such self-handicapping behavior is also called emotion-focused, dysfunctional self-regulation (Steel, 2007).

With regard to virtual worlds, goals and tasks are an inherent part of their design (Salen & Zimmerman, 2004). Even if solely fun-oriented, it is difficult to find a virtual world which does not include some kind of feedback on achievement (related to its specific content) as a permanent feature. Self-motivation should thus have the potential to further explain IR performance.

In line with previous research–and in contrast to intrinsic motivation being conceptualized as computer playfulness closely related to a particular application field (Venkatesh, 2000)–we suggest to conceptualize self-motivation as a personal difference unrelated to a specific task or activity. We refer to motivational skills as personal traits (Kanfer, Ackerman, & Heggestad, 1996), that is, core attributes of an individual which have broad implications for behavior and affect. However, one needs to be aware that self-report measures of motivational and self-regulation skills may relate to certain personality traits (Kanfer et al., 1996; Devaraj et al., 2008). The capability of self-regulation, thus the degree to which the examined individual is capable of self-regulation for motivation (Wong & Law, 2002), is an important quality to meet requirements which have been set. Additionally, in accordance with the definition of self-motivational traits, every individual has set certain–quantifiable–general and intrinsic standards for himself or herself (Frost, Marten, Lahart, & Rosenblate, 1990) and tries to live up to these standards.

h7 Self-motivation capability is positively related to IR performance.

h8 Intrinsic goal orientation is positively related to IR performance.

This conceptualization accounts for the aforementioned dimensions in the form of a first-order construct.

COMPETITIVENESS Competitive goals, or performance goals, respectively, are linked to an individual's goal of establishing superiority over others (Yperen & Janssen, 2002). Competitive excellence, that is, the desire to compete with and perform better than others, is an important approach motivation (in contrast to avoidance motivation) related to workplace (Diefendorff & Mehta, 2007). Murayama and Elliot (2012) stated in their meta-analysis that in the past, competition has been conceptualized as "a characteristic of a person (*trait competitiveness*), as a characteristic of the perceived situation (*perceived environmental competitiveness*), and as a characteristic of the actual situation (*structural competition*)... " (p. 1035, emphasis in original). All three approaches capture interpersonal competition (i. e., competition between individuals) as opposed to intrapersonal competition (i. e., competition with oneself) and intergroup competition (i. e., competition between groups), the latter two of which have not received the same attention as the former in the past (Murayama & Elliot, 2012). Moreover, management science has generated many theories to explain oliogopolistic competition. However, these theories have often only been tested through simulation games (Remus, 1978).

In the present study we focus on competitiveness as a personal characteristic, as both situational and structural settings were (a) difficult to account for and (b) considered to be highly competitive in the context of study. We captured competitiveness through a person's seeking for superiority (Yperen & Janssen, 2002) and added a similar aspect, namely perfectionism, to tap into a wider range of dimensions of what we assumed constituted the driving force of an individual to engage in competition. To achieve this, we included the tendency to strive for perfection and the level of personal standards of an individual (Frost et al., 1990)–both aspects representing the (positive) dimensions of perfectionism (Stoeber, Stoll, Pescheck, & Otto, 2008)–comparing oneself with others in a general fashion. Following Stoeber et al. (2008), we are thereby in line with the contemporary approach to achievement orientation which focuses on goals and investigates the different reasons why individuals are eager to achieve.

While the content and purpose of some virtual worlds mainly appeals to the social needs of their users, others foster competition (Klimmt, Schmid, & Orthmann, 2009; Weiss & Schiele, 2013). Certain types of virtual worlds are essentially "motivated by the desire to gain power and progress rapidly in the game" (Frostling-Henningsson, 2009, p. 559), thus to aspects of competitiveness in relation to performance.

h9 Seeking superiority is positively related to IR performance.

h10 Striving for perfection is positively related to IR performance.

h11 Personal standards in comparison with others is positively related to IR performance.

We consider competitiveness a three-factor construct, possibly in the form of a higher-order structure.

IR ENJOYMENT Heijden (2004) defined *perceived* enjoyment as the extent to which fun can be derived while using a specific system. Similarly, Agarwal and Karahanna (2000) defined *heightened* enjoyment as a subdimension of the state of cognitive absorption which captures the "pleasurable aspects of the interaction" (p. 673). Both concepts describe a situational component of an experience, in contrast to playfulness (Agarwal & Karahanna, 2000), microcomputer playfulness (Webster & Martocchio, 1992), and computer playfulness (Venkatesh, 2000), which all represent traits and which do not depend on a particular system. Wakefield and Whitten (2006) yet argued that enjoyment is an affective construct. They criticized the mixing of (heightened) enjoyment with the dimensions of control, curiosity, immersion and temporal dissociation and aggregating them into only one construct (namely cognitive absorption, Agarwal & Karahanna, 2000), because in their view, the latter four dimensions rather represent cognitive functions.

Following the arguments above, we decided in favor of a situational construct, also because the scope of the general trait of playfulness is too broad for our purposes and difficult to account for. Additionally, we assumed (micro)computer playfulness to be high for *all* our participants, as the latter engage in extensive computer gaming. It therefore made more sense to capture the pleasure and enjoyment of a specific IR system (like a virtual world) on an individual to gain further insights into the interaction between system and system usage.

One of the main reasons for using virtual worlds is their support for fun activities through social interaction (Hundley & Shyles, 2010). Research in the gaming context suggests that intrinsic motivation in pursuing a particular activity–as opposed to extrinsic motivations–leads individuals to process information more carefully and completely, making them exhibit higher levels of creativity and cognitive flexibility (Przybylski et al., 2010)–all of which are factors which should contribute to higher IR performance.

h12 IR enjoyment is positively related to IR performance.

IR MEDIATION The effects of CIRME may be amplified if an individual receives active mediation or guidance by another individual. In the context of television, so-called instructive, evaluative, or active parental mediation goes beyond simple coviewing (Gentile et al., 2012) or rule making (Fujioka & Austin, 2002), as it involves conversations about the medium and its content. To this end, parents may point out good and bad things on TV or try to explain things that happen and whether they are realistic or not (Valkenburg, Krcmar, Peeters, & Marseille, 1999). Hence, this type of mediation helps children to (a) categorize and (b) validate messages related to television, but also to (c) gain supplemental information on the "potential relevance of television messages to real life" (Fujioka & Austin, 2002, p. 645).

Based on this definition (cf. Nikken, 2003), IR mediation relates to the extent to which an individual has experienced mediation by another person at the age of 10 to 14 with regard to an IR medium. Our definition is yet not limited to mediation by parents, thereby accounting for the fact that siblings and peers have an important influence on a child's behavior (A. B. Kelly et al., 2011) and that siblings provide one of the most stable and powerful contexts for the social-emotional development of a child (Stormshak et al., 2009).

With regard to virtual worlds, individuals who participated in virtual reality trainings which included perceptual training performed better when they had been given the opportunity to receive instructions containing information on expert performance strategies (A. M. Williams et al., 2002).

h13 IR mediation positively affects the influence of CIRME on IR emotion capabilities such that its effect is stronger for individuals which have experienced more IR mediation.

As mediation estimates reported by children and parents often differ (Buijzen, Rozendaal, Moorman, & Tanis, 2008) and because parental self-reports may reflect currently accepted parenting behaviors rather than parents' actual past activities in terms of mediation (Gentile et al., 2012), we relied on mediation experience reported by our participants. No distinction was thereby made between the different mediating actors (e. g., relative or friend), as we suggest that any mediation would have added to the media understanding of a child. We also assume this construct to be unidimensional.

PARENTAL CONTROL Parental control refers to the degree to which parents have monitored and restricted an individual's IR media use during its childhood. Activities to control media use comprise (a) checking the contents of a medium while and after consumption, and (b) limiting the consumption to certain times and to a certain amount (cf. Valkenburg et al., 1999; Valcke, Bonte, Wever, & Rots, 2010; Gentile et al., 2012). These aspects do not represent distinct dimensions of the construct, but rather facets of a more or less consistent parenting style. It can be assumed that this type of parenting style may interfere with the learning process related to IR emotions, for example, if a child needs to abruptly leave an IR setting due to media restriction rules before being able to fully analyze a particular situation. Parental control should thus negatively affect the impact of CIRME in such a way that individuals which have experienced stricter control may have been interrupted during the development of IR emotional capabilities; hence they may not have harvested the full potential of their IR experiences.

h14 Parental control negatively affects the influence of CIRME on IR emotion capabilities such that its effect is smaller for individuals which have experienced more parental control.

Studies suggest that a child's report on parental monitoring may be a more reliable source than a parent's report, due to effects of social desirability (Gentile et al., 2012). We therefore chose to rely on the "former" children, that is, our now grown-up study participants, for an evaluation of the control carried out by their parents at the age of 10 to 14 with regard to IR media.

EXPERTISE Experiments implementing video-based tests of anticipation in tennis have shown that skillfully anticipating of others' actions–like the direction of opponents' tennis strokes–develops with a player's expertise. Also, pure reaction time did not account for the effects of tennis skill level (Rowe & McKenna, 2001). Results further indicated that more effective visual search behaviors can actually make up for less speed due to age, for example (A. M. Williams et al., 2002). These results are transferable to virtual worlds.

h15 Hours of play positively affects a user's skilled anticipation of other participants and thus his or her IR performance.

MOTIVES As discussed above, various factors in terms of self-motivation and competitiveness are potentially relevant regarding IR performance. Whether a virtual world user is purely interested in leisure aspects of virtual world usage, or whether the user is a professional using the virtual world in order to make a living instead, should have an impact related thereto. We therefore consider the motives for play an important control variable.

h16 Motives for play vary across groups in that levels of self-motivational traits, competitiveness, and enjoyment may differ in the distinctive groups such that the motives for play moderate the effects of the latter constructs.

This chapter has discussed prior research, depicted our initial model, presented our construct conceptualizations, and illustrated the development of our primary and secondary hypotheses. The upcoming chapter treats the development of measurements and the conduction of the pretests as well as the conduction of the actual study.

Table 2: Dimensionality of primary and secondary constructs as hypothesized by their respective conceptualizations

Construct	Factors	Hypothesized dimensions
IR Performance	1	Unidimensional
Perceive IR Emotions	1	Unidimensional
Understand IR Emotions	1	Unidimensional
Manage IR Emotions	1	Unidimensional
SE Cognitive Ability	1	Unidimensional
Cognitive Absorption	2	(a) Focused immersion and (b) temporal dissociation
CIRME	4	(a) Watch IR action passively, (b) experience IR on one's own and alone, (c) experience IR with others via network, and (d) experience IR with others in physical proximity
Self-Motivational Traits	2	(a) Capability of self-motivation and (b) intrinsic goal orientation
Competitiveness	3	(a) Seeking superiority, (b) striving for perfection, and (c) personal standards in comparison with others
IR Enjoyment	1	Unidimensional
IR Mediation	1	Unidimensional
Parental Control	1	Unidimensional

3 CONTEXT, CONSTRUCTS, MEASUREMENTS, AND DATA COLLECTION

With insufficient data it is easy to go wrong.

— *Carl Edward Sagan*

We have outlined our construct conceptualizations, the development of hypotheses, and the resulting structural model in the previous chapter. This chapter now begins with a description of the general context of the study and its participants. It then discusses the assumed type of relationships between the model constructs and their respective indicators, gives details about the measurement development process, and presents the measurements finally used. It then outlines the course of the pilot studies as well as of the actual data collection.

3.1 CONTEXT AND PARTICIPANTS OF THE STUDY

Our study was conducted in the context of competitive online computer games, often generically summed up under the term eSports. We examined one of the largest eSports community in Germany, the latter of which is formed by a particular group of users who are members of a certain specialized eSports portal on the Internet. At the time of our study, the community had approximately 320,000 members,[1] out of which approximately 53,000 were admitted to our questionnaire. The company that operates the portal has its registered office in a German city, but access to its services is not restricted to Germany: The portal has about 4,000,000 registered members in 41 countries,[1] organized in national and transnational communities. In many countries, the aforementioned portal operator is the organizer of the national eSports league which elects the national eSports champions in several categories. To support activities abroad, the operator has set up establishments in North America, France, Italy, Poland, Spain and China.[1]

Among all members of the portal, the German members constitute one of the largest subgroups and form a community of their own. Users are thereby not assigned to the "German" community according to their nationality or language, but according to their geographical location. More precisely, only users with a German Internet protocol (IP) address are admitted. General admission to the portal is generally open to everyone and free of charge. However, to benefit from certain amenities (no banner ads, access to exclusive leagues, etc.), a user needs to get a premium account and pay a small monthly fee. Moreover, as high

1 Information from the operator's company brochure and/or web site.

gaming rankings are associated with high prestige, world-wide visibility, and high value prizes, the level of cheating prevention enforced by the operator increases with the gaming level. Championship contenders are therefore requested to undergo well-defined procedures to (a) prove their identity and (b) demonstrate compliance with regard to technical requirements and the system state of the computer they use, for the purpose of being granted a certain "trusted player" level and thereby gain or continue to have access to certain leagues.

While visiting the community portal, users can find and challenge contestants and take part in leagues of more than 130 different games. Analogous to nondigital sports like soccer or tennis, either single athletes[2] or teams compete against each other. There are seasons, rounds, matches and finals, with different types of tournament elimination and scoring systems. The community can roughly be divided into casual, semi-professional, and professional gamers. Real-life eSports events take place regularly, and spectators can either watch them on-site or via television broadcast and Internet stream, respectively. In case of conflict with a contestant, technical problems, or other issues, users receive support from chosen volunteers of the community or by the operator's staff. The portal also serves as a source of social support, as features of the platform allow for chatting, exchanging news, and arranging meetings. Services provided by the operator include hosting the league matches, organizing the rankings, cheat-prevention (e. g., through special software), and monitoring the discussions in the forums.

Our final questionnaire was part of the portal operator's biennial survey which is sent to all registered members of that community. To ensure comparability and to account for the fact that cultural differences have been identified as important moderators, for instance of technology acceptance (cf. Srite & Karahanna, 2006), we excluded German-speaking communities of other countries from our study (see also Likert, 1932, on this topic). For more details on the sampling processes for the pilot studies and the actual study, the reader is referred to Section 3.3 and Section 3.4, respectively. An overview on the exact numbers of participants is given in Table 7 at end of this chapter.

3.2 MEASUREMENT DEVELOPMENT

After having defined our constructs and their dimensionality, we needed to "step back and evaluate (...) how (...) [the multiple subdimensions of the focal construct] relate to the focal construct and to each other" (MacKenzie et al., 2011, p. 300) before deciding on adequate measurements.

2 Professional eSports player are officially recognized as professional athletes by the U.S. government, see also http://www.forbes.com/sites/insertcoin/2013/07/14/the-u-s-now-recognizes-esports-players-as-professional-athletes/.

3.2.1 *Relationships Between Factors and Indicators*

Hereinafter, we shall relate to the measurement model as the part of a model "which shows the relationships between observed variables that are indicators of latent variables, and the latent factors they represent", and to the structural model as the part "which shows the relationships among the latent variables in the measurement model, and the observed variables that are not indicators of any of these latent variables" (M.-h. R. Ho, Stark, & Chernyshenko, 2012, p. 44, see also Gefen et al., 2011).[3] We make this explicit distinction because authors sometimes talk about *the* measurement model when what they really refer to is a certain quality of the measurement model. In such case, researchers essentially qualitatively assess the nature of relationship which may exist between the latent variables (i. e., constructs, but also factors or subdimensions, respectively) and their indicators, and which should be reflected by the directionality of their association (for an excellent overview and clarification of terms, see Bollen, 2011).[4] We recap the ongoing debate on when to use which type of measurement model as well as the possible consequences of misspecification later in Table 5.3 again.

This section deals with how we determined what type of measurement model was appropriate for our constructs. Our approach concerning this issue was mostly based on the recommendations of Jarvis et al. (2003) and consisted of (a) rereading the conceptualizations of the original works we had taken into account for our definitions, (b) reflecting on their possible factor-indicator relationships, and if necessary or applicable, (c) comparing the factor-indicator relationships we hypothesized to those of the same or a similar construct of other authors, (d) compare our constructs to the constructs to find matching and explicit recommendations on the subject as listed, for example, in publications cited above, and (e) considering arguments for and against one type of relationship direction and the respective alternative (i. e., reflective vs. formative and vice-versa) based on our definition of the construct.

We illustrate the procedure using the example of cognitive absorption. The latter has its roots in IS and represents one of its genuine concepts; it is therefore very well suited for this purpose because most readers will be familiar with the basic notion on what it captures. As noted before, it has been introduced as a reflective construct by Agarwal and Karahanna (2000) and has been reimplemented by Burton-Jones and Straub (2006) and Goel et al. (2011) in the same manner. Bearing in mind that our conceptualization of the construct focuses on temporal dissociation and focused immersion, we applied the decision rules de-

---

3 In PLS the two parts are typically referred to as the inner and outer model instead, compare Wetzels et al. (2009).

4 Note that latent variables can be indicators of higher-order latent variables themselves.

veloped by Jarvis et al. (2003) and integrated into our aforementioned procedure as follows:

1. We argue that when reasoning about the direction of causality, both dimensions are manifestation of the focal construct and that a change in the construct should produce a change in all of its subdimensions. A person experiencing a higher level of cognitive absorption should perceive higher levels of temporal dissociation *and* focused immersion, and both reflect the level of cognitive absorption. Consequently, cognitive absorption exists separately at a deeper level than its subdimensions (cf. MacKenzie et al., 2011).

2. As stated by Burton-Jones and Straub (2006) and repeated by Goel et al. (2011), the indicators can be viewed as interchangeable. Also, they share a common theme, and dropping one of them (as has been done in both studies) should not alter the conceptual domain.

3. The chosen indicators for each of the factors or subdimensions (temporal dissociation and focused immersion) are certainly expected to covary. Moreover, we find it difficult to envision a user who is highly absorbed in interaction with a system whose temporal dissociation and focused immersion point into different directions.

4. In our view, temporal dissociation and focused immersion should not have different antecedents or consequences, nor should their respective indicators.

Furthermore, to the best of our knowledge, cognitive absorption has no other "historical" record, nor has a similar construct in literature received a differing recommendation concerning the type of measurement model. It therefore seemed appropriate to model cognitive absorption as a reflective construct–at the construct level as well as at a potential factor level (in case of support for a higher-order structure).

This process, as shown by example above, was applied to all constructs; if a construct was assumed to have several underlying factors (subdimensions), the procedure was adjusted accordingly. After careful consideration, we concluded that a reflective measurement model was suitable for all our constructs, respective factors (where applicable), and their suggested indicators.

3.2.2 *Measurements for Constructs and Controls*

On the basis of our literature review, our model construct conceptualizations, and the postulated factor-indicator relationships (determined in the previous section), we generated an initial list of candidate items.

The majority of items we used were adapted from previously published measurements; an overview of the measurement sources is shown in Table 3 and Table 4. Following suit with many examples of reflective measurement models in literature, we opted for multiple items measurements (Weiber & Mühlhaus, 2010). In the case of a preexisting scale, the amount of indicators was thereby generally given be the amount of items the scale consisted of, whereas in the case of a newly developed scale, we followed the advice to use at least two, better three or more indicators per factor for a model "with two or more reflective factors" (Bagozzi, 2011, p. 271; see also Weiber & Mühlhaus, 2010, p. 93 et seqq.).

Most existing scales and their corresponding items were in English, as they originated from articles published in international journals, yet to avoid any language problems with regard to the participants and to align our questionnaire with the survey of the operator, it was necessary to translate English items into German. We also slightly adjusted items to the context of the study where necessary. Other variables were newly developed, like specific items on childhood events to countercheck responses, for example, or others intended to gain a better understanding of our target group. The final items as well as the original English items (if applicable) can be found in Appendix A.

In parallel to our efforts to develop measurements, the operator's staff developed items of their own. These items were of particular interest to the operator and were to be added to the same survey at the end of our questionnaire. The operator items requested survey candidates to specify details on their behavior as a consumer and on their satisfaction with the platform. The process of adding items to our survey was repeated before every data collection wave except Pilot 2 (see Section 3.3). As to avoid redundancy and because items on social demographics and some of the questions on motives for playing were also interesting to us, we coordinated our part with that of the operator; as a result, an item of our survey can either a) be attributed to our study, or b) was of interest to both the operator and our study, or c) was only of value to the operator. The items of all data collection waves (as already mentioned, with the exception of Pilot 2) can be classified as follows:

- items of the model constructs,
- items designed for control purposes,
- items on gaming preferences, motives for playing, awareness of cheating, and requests for improvement suggestions,
- items on consumption habits and on satisfaction with the portal and the leagues,
- items on demographic characteristics,

- items of self-disclosure, and
- a free text comment.

In principle, the enumeration above mirrors the order in which the different item types appeared in the pilots as well as in the actual study, with the important exception that we ensured our own age item was placed as the second item of the whole survey, to be able to assign age groups and possible differences between them even in case of dropout.

The following sections present the measures used for the study in a detailed manner, first those of the model constructs, divided into primary and secondary constructs, followed by the measures of all other variables of interest. With the exception of performance, all items were measured through bipolar rating scales–either agreement or intensity ratings–using a seven-point response format, for which each data point was provided with a combined verbal and numerical label in the questionnaire. The anchors for agreement ratings ranged from “strongly disagree (-3)” to “strongly agree (+3)”; intensity ratings were more diverse in terms of wording for suitable meaning (e. g., “a lot more than my classmates (-3)” to “a lot less than my classmates (+3)”).

3.2.2.1 *Measurements for Primary Constructs*

Construct conceptualizations were presented in Section 2.5, and justification for using a reflective measurement model was given in Section 3.2.1. The following section presents the measurement details of our primary constructs; a short summary can be found in Table 3.

Performance was a special case in terms of measurement, as it was not assessed via questionnaire. We measured this construct via the user’s player level, a single and manifest variable, hence no scales or items were developed for this measure. The score is computed by the internal league system of the portal, which generates it for every user and displays it on the user’s profile page automatically. It is computed as follows: For every single game a user plays via the portal, an individual game level is computed, reflecting the user’s achievements in that particular game in comparison with every user who plays this very game in the community. All game levels of a user added up–that is, the sum of a user’s game levels–reflect the overall player level a user. A user’s player level is updated by the system when one of his or her game level changes. We retrieved this information with the help of a web crawler which gathered this information from the profile pages of our survey candidates via Internet. Technical details on the data retrieval are given in Section 3.4.1.2. Though this item was measured continuously and seemingly obeyed the rules of a ratio scale, assuming an interval scale of measurement for this item seemed appropriate as well. This depended, among other things, on whether, as an example, a score of two (or eight) could be meaningfully be interpreted as the double of score on (or four, respectively); the final decision was to be made based on the final data distributions.

Table 3: Sources of measurements for primary constructs

IR Performance	Single item measure based on a user's player level, an objective unidimensional measure of performance; it represents the overall achievement of the user in all games taken together.
IR Emotional Capabilities	Items adapted from TEIQUE 1.5 and TEIQUE-SF, Wong and Law (2002), E. J. Austin, Saklofske, Huang, and McKenney (2004), and Tett, Fox, and Wang (2005) measuring the dimensions (a) emotion perception, (b) emotion understanding, and (c) emotion management; few new items targeting similar but more specific aspects of virtual worlds and eSports.
SE Cognitive Ability	New scale measuring an individual's astuteness, comprehension and learning speed, capacities compared with the average, educational background, intellectual giftedness, and academic achievements from the viewpoint of others, estimated by the individual investigated.
Cognitive Absorption	Items adapted from Agarwal and Karahanna (2000) measuring the dimensions (a) focused immersion and (b) temporal dissociation.
CIRME	New scale measuring the childhood IR media experience and the social setting of that experience along different dimensions, that is, (a) passively, (b) alone, (c) with others via network, (d) with others in physical proximity, estimated in retrospective.

Note. Names of the scales for each dimension here correspond to the names given in the original work and do not necessarily correspond to the name of the construct dimension they were labeled with in this study.

As aforementioned, we implemented the cascading model postulated by results of the meta-analysis by Joseph and Newman (2010). The latter tested their model empirically using a correlation matrix which they constructed from a large number meta-analytic estimates. We instead performed a self-contained analysis on the basis of self-reports, as suggested by other studies on the subject (e. g., Lane et al., 2009). We measured IR emotional capabilities as three distinct constructs–

perceiving, understanding, and managing IR emotions (cf. Joseph & Newman, 2010)–using selected items of the Trait Emotional Intelligence Questionnaire (TEIQUE) version 1.5, and its short form, the TEIQUE-SF (cf. e. g., Freudenthaler et al., 2008; Cooper & Petrides, 2010) as published by Tett et al. (2005), E. J. Austin et al. (2004), and Wong and Law (2002). Items were selected in such a way that they would cover the capability facets of emotion perception, emotion understanding, and emotion management (cf. Joseph & Newman, 2010; Brackett et al., 2006) with regard to performance (e. g., use of emotions to facilitate performance, Law et al., 2004), and then split into our three postulated constructs (see Table 2 on p. 74). If needed, own items with similar wordings were developed that were specifically related to virtual worlds and eSports.

With regard to cognitive ability, psychology has developed a large variety of theories and many measurement instruments in accordance with them. Examples are the Culture Fair Intelligence Test (CFIT), which tries to reduce the effect of cultural or educational background to a minimum (Cattell, 1940), and the Wonderlic Personnel Test (WPT).[5] The former has appeared in different versions and contains several subscales (e. g., Cattell, 1973), the latter is designed for employee selection, training, and placement. Côté and Miners (2006) used them to evaluate the outcome emotional and cognitive intelligence on job performance in the context of administrative sciences. The WPT also served to assess the general mental ability of U.S. NFL players to predict their NFL performance (Lyons et al., 2009) as well as to investigate interactions between on self-efficacy, performance, and knowledge (B. S. Bell & Kozlowski, 2002), and it was used it to control for cognitive ability to examine the success of computer training methods by S. J. Simon et al. (1996). We developed measures based on the characteristics of the WPT reported by the latter and additionally accounted for cultural background, experience, and education (cf. Furnham et al., 2005); for the above reasons, these measures were specifically designed to measure SE cognitive ability.

Cognitive absorption was measured using items developed by Agarwal and Karahanna (2000), accounting for the dimensions of temporal dissociation and focused immersion.

Childhood experience with IR media was measured through measurements we developed that were based on many studies accounting for the amount of time spent in front of the investigated medium (D. R. Anderson & Hanson, 2010); we altered them by giving them a more social perspective on media use (Haefner & Wartella, 1987; Gentile et al., 2012; Bers, 2010). This new scale accounted for self-estimated CIRME and for the extend this experience was shared with others (transmitted via the medium or directly).

5 http://www.wonderlic.com/

3.2.2.2 *Measurements for Secondary Constructs, Controls, and Other Special Items*

This section first presents the measurement details (summarized in Table 4) of constructs additional to our initial model referred to as the secondary constructs, followed by the presentation of control items and other special items which, as for instance, accounted for possible subgroups among our participants.

Self-motivation made use of personal standards items (Frost et al., 1990), but only with regard to oneself, and additionally included the use of emotion for self-motivation (Wong & Law, 2002). The three-dimensional construct of competitiveness, in turn, included the dimensions of performance orientation, strive for perfection, and personal standards, too, yet comparing oneself to others and was measured using items developed by Yperen and Janssen (2002), Stoeber et al. (2008), and Frost et al. (1990). Enjoyment while using a particular IR system was measured using items of different enjoyment scales developed by Agarwal and Karahanna (2000), Heijden (2004), and W. W. Chin and Lee (2000). Mediation between the real world and the virtual world, that is, IR mediation, was measured unidimensionally using items of Nikken (2003) and Valkenburg et al. (1999), and we used parental control measures developed by Valcke et al. (2010) and Valkenburg et al. (1999) to capture the unidimensional construct of parental control.

Furthermore, we developed several items to control for consistency of responses with regard to childhood experience items, to ensure answers would be valid with regard to time periods and the occurrence of events. Their use seemed helpful after experiences with the pilot studies. Other items were developed to complete the picture of a user's gaming behavior, gaming preferences, and awareness of cheating. Similarly, other items gave users the opportunity to improve the leagues according to their preferences; the latter items were designed to gain a better understanding of our target group, but also to encourage the participants to finish the survey. An example of this type of items were requests for improvement suggestions on topics such as in-game emotion perception, the leagues, the forum etc., so questions like "Would you be interested in more means to perceive the emotions of your opponent?" or "What game genre should the league pay more attention to in the future?". Moreover, we aimed at controling for cultural differences due to a migration background. We thus ask participants to specify whether they were born in Germany, whether German was the principally spoken language at home, and whether they actually lived in Germany (as IP addresses do not necessarily reflect the actual location of a user), and we also inquired about the educational background of the parents.

Table 4: Sources of measurement scales for secondary constructs

Self-Motivational Traits	Items adapted from Wong and Law (2002), use of emotion (UOE); Frost, Marten, Lahart, and Rosenblate (1990), (general) personal standards.
Competitiveness	Items adapted from Yperen and Janssen (2002), performance orientation; Stoeber, Stoll, Pescheck, and Otto (2008), striving for perfection; Frost, Marten, Lahart, and Rosenblate (1990), personal standards (with regard to others).
IR Enjoyment	Items adapted from Agarwal and Karahanna (2000), heightened enjoyment; Heijden (2004), excitement; W. W. Chin and Lee (2000), overall satisfaction (item set 1).
IR Mediation	Items adapted from Nikken (2003), evaluative mediation; Valkenburg, Krcmar, Peeters, and Marseille (1999), instructive mediation.
Parental Control	Items adapted from Valcke, Bonte, Wever, & Rots, 2010, supervision, Internet usage rules, and stopping Internet usage; Valkenburg, Krcmar, Peeters, & Marseille, 1999, restrictive mediation.

Note. Names of the scales for each dimension here correspond to the names given in the original work and do not necessarily correspond to the name of the construct dimension they were labeled with in this study.

3.2.2.3 *Shared and Operator Items*

This section describes the items the operator developed for the purpose of collecting information on the survey candidates as consumers and users of the portal; the reader is referred to item 5.2 for additional information on this subject.

ITEMS OF SHARED INTEREST Items of this category recorded demographic details on the survey candidate like age, gender, occupation, the state (i. e., the Bundesland) they lived in, the number of inhabitants of their city, and their educational background. The survey candidate was also requested to specify his or her preferred game genre. To understand the motives for playing games, a user was additionally asked to rate how serious he or she took the leagues, how important it was to win prizes, and whether fun or rather money was the main driver for playing. The questionnaire ended with items of self-disclosure which (a) asked whether the survey candidate had answered the questions with care and faithfully, or whether the candidate had only quickly clicked through the survey in order to take part in the raffle; furthermore, again as a potential means to check the validity of answers, the user was (b) requested to rate the length of the study. At the very end, the survey candidate could leave a free comment.

CONSUMER BEHAVIOR AND SATISFACTION Questions on aspects of consumer behavior sampled information on, for example, what type of technical devices could be found in the users' households, how they would usually become aware of new technical devices, or which type of products they would buy online. Questions on satisfaction asked for their opinion on certain portal features, reasons for willing or not willing to pay the monthly fee associated with a premium account, and so on. Detailed numbers and assignment of items to different categories in the pilots are shown in Section 3.3.2, whereas Section 3.4 presents the same information for the actual study.

Another category of items which is not explicitly listed in the tables are the 10 variables stored by LimeSurvey,[6] the open source application used to implement the surveys. These variables are automatically generated and serve documentation purposes (like, e. g., start and end time and date), yet are invisible to survey candidates.

3.2.3 *Assessment of Content Validity*

Though most of the items we intended to use belonged to existing scales, we needed to avoid translation or language issues and to validate our "questions by both experts in the field and laypeople" (Allport & Kerler, 2003, p. 356). During the whole measurement development process as well as through all four data collection waves, we were supported by a certain employee of the portal operator. He was responsible for the conduction of the survey and particularly designated to support our project. In that capacity he was available to all our questions, and his insights into the community structure, social demographics, and internal matters provided guidance and important impulses. He himself was not only a member of the operator's staff, but also involved as a user of the portal, thus we profited from the expertise of a subject matter expert *and* practitioner at the same time. With his support and the help of the other staff members he had won for our project, we were able to assess the adequacy and representativeness of our measures–especially with regard to the wording of items–in several sessions, even before the beginning of the pilot studies (cf. e. g., MacKenzie et al., 2011).

6 https://www.limesurvey.org/

3.3 PILOT STUDIES AND PRETESTS

Collecting data via pilot studies to conduct pretests has been recommended by many authors as an important step during measurement construction (e. g., MacKenzie et al., 2011; Weiber & Mühlhaus, 2010; Straub, 1989; Churchill, 1979). Though most of our items were part of existing scales which were validated and published in their original sources, we decided to pretest all our items as if they were newly developed, that is, all items that had "survived the content validity assessment" (Hinkin, 1998, p. 110).

3.3.1 *Setup and Implementation*

The first pilot study we prepared, referred to as Pilot 1 hereinafter, covered all items we had developed so far. Pilot 1 was launched end of January 2011 and lasted two weeks. As with all following surveys, its implementation was supported by LimeSurvey, the aforementioned survey tool. The participants of Pilot 1 were the members of an exclusive discussion forum called the moderator forum, a virtual space admitted only to so-called admins of the portal. Admins are members of the community and volunteers willing to take on tasks like moderating discussion threads, giving technical support as well as mediating between opponents of a challenge in case of disputes, all without payment. Their activities at the intersection between the community and the operator are crucial to the portal, as they ensure smooth operation in terms of the daily league business. They use the moderator forum to discuss issues of league management, cheating problems, software releases, and other topics related to eSports and the portal. During the course of this pilot study, 69 cases were collected; the survey tool system reported that for approximately half of them, the questionnaire was incomplete.

To run pretests on more data, we then implemented two further pilot studies, Pilot 2 and Pilot 3. They were launched simultaneously in March and ended in April 2011. The questionnaires of both pilot studies were slightly adapted compared with Pilot 1 if results of data investigations implied this. During the course of Pilot 2 and Pilot 3, 180 (90 reported incomplete by the system) and 590 (118 reported incomplete) cases, respectively, were collected. Pilot 2 only covered our self-developed items on CIRME, while Pilot 3 covered all candidate items that had been developed so far (like Pilot 1 before). An overview on the different pilots and their structure can be found in the next section.

3.3.2 *Purification and Refinement of Measures*

To obtain purified scales, we examined their reliability based on Cronbach's α and item-scale statistics on the indicator and scale and/or construct level, respectively (MacKenzie et al., 2011; Bühner, 2011; Weiber & Mühlhaus, 2010; Gignac, 2013), thereby using listwise deletion in case of missings (cf. Roth, 1994). Additionally, to discover how many components best described our data (cf. Straub, 1989), we conducted a principal component analysis (PCA). PCA was thereby applied *after* the purification process, to avoid producing "many more dimensions than can be conceptually identified" (Churchill, 1979, p. 69).

According to observations we made and remarks we received by participants or the operator staff, we eliminated or rephrased several items, for instance if their wording appeared to be misleading and the intended meaning had not become clear. This process is documented in Table 5. Details on the number of the remaining items in the final survey are shown in Section 3.4, and the actual items can be found in Appendix A.

3.4 ACTUAL DATA COLLECTION

After finalizing our measurements, we collected the actual data of the present study from a new sample. Again, the collection of survey data was supported by LimeSurvey.

3.4.1 *Means of Gathering Data*

Data were collected using two different means. The first means was through questionnaire items, which resulted from the measurement development. In total, the file that contained the results of our actual survey enclosed 6,779 cases (3,417 reported incomplete by the system); Table 6 illustrates the survey structure.

The second means consisted of gathering information on participants' individual gaming performance from their individual profile pages. Data were gathered via web crawler (see Section 3.4.1.2 for more details). We thereby retrieved 6,616 data points which were later matched with the survey entries via the unique identifier of a player, the so-called player ID. An individual player ID is assigned to every user during the initial registration process as a unique identifier, and as such, the player ID represents an integer from a consecutive series of numbers, starting at 1 (for more details, see upcoming sections).

3.4.1.1 *Collection of Survey Data via Questionnaire*

Our final questionnaire was part of the portal operator's biennial survey sent to all registered members of the German community. The survey was, like in the years before, divided into several different self-contained questionnaires. Overall six different questionnaires including ours were part of the survey 2011. The whole survey with all its questionnaires was accessible for a period of four weeks, from beginning of May until beginning of June 2011, and hosted on the operator's server.

Table 5: Number and types of items used in the pilot studies

		Assigned to[a]			
Wave	Item type	St.	Sh.	Op.	Σ
Pilot 1					
	Constructs & controls[b]	139	0	0	139
	Preferences/motives[c]	12	4	0	16
	Consumer behavior[d]	0	0	15	15
	Demography	4	8	1	13
	Disclosure[e] & comment	0	3	0	3
	Total	155	15	16	186
Pilot 2[f]					
	Constructs & controls[b]	67	0	0	67
	Preferences/motives[c]	2	0	0	2
	Comment	1	0	0	1
	Total	70	0	0	70
Pilot 3					
	Constructs & controls[b]	171	0	0	171
	Preferences/motives[c]	12	4	0	12
	Consumer behavior[c]	0	0	18	18
	Demography	4	7	1	12
	Disclosure[e] & comment	0	3	0	3
	Total	187	14	19	220

[a] Items of category (i) St. = study, (ii) Sh. = shared, (iii) Op. = operator.
[b] Items to countercheck validity of answers.
[c] Preferences, motives for playing, awareness of cheating, and request for improvement suggestions.
[d] Questions on buying behavior and satisfaction with the league.
[e] Gave survey candidates the opportunity to disclose whether they had answered the questions honestly and to rate the length of the survey.
[f] Only items for the CIRME construct were sampled.

As with Pilot 1 and Pilot 3, the operator added items to our final questionnaire, too, so that various topics of interest to the operator (e. g., on how to improve the leagues, a user's planned purchases and lifestyle habits, detailed questions on particular games, etc.) were also covered by it. After all questionnaires were implemented, their launch was announced in the portal news. Additionally, all active members[7] of the German community received an invitation email containing a direct link to the survey. The invitation email, the news text as well as the introductory text to the actual survey informed a survey candidate of the cooperation with us. The introductory text also stated that participating in the survey could help to explore the phenomenon of eSports further and reduce prejudices against gaming in general. It further contained the warranty that any information given would be treated as strictly confidential and that anonymity was kept at all times. To encourage participation, the operator raffled 100 vouchers for a popular music download portal, each either worth €10 or €20, accounting for a total of €1,400 (for a discussion on lottery incentives and response rates, the reader is referred to Section 5.2). Below the abovementioned information, we briefly presented our university and department and said a few words about the abstract goals of our study–gaining insights into a user's experiences and interpret them with regard to the use of virtual worlds in work contexts. We also mentioned that our questionnaire would be used to examine how to contribute to fairer leagues.

Table 6: Number and types of items used in the actual study

	Assigned to[a]			
Item type	St.	Sh.	Op.	Σ
Constructs & controls[b]	108	0	0	108
Preferences/motives[c]	13	4	0	17
Consumer behavior[d]	0	0	18	18
Demography	5	8	1	14
Disclosure[e] & comment	0	3	0	3
Total	126	15	19	160

[a] Items of category (i) St. = study, (ii) Sh. = shared, (iii) Op. = operator.
[b] Items to countercheck validity of answers.
[c] Preferences, motives for playing, awareness of cheating, and request for improvement suggestions.
[d] Questions on buying behavior and satisfaction with the league.
[e] Gave survey candidates the opportunity to disclose whether they had answered the questions honestly and to rate the length of the survey.

A questionnaire had randomly been given a number from 1 to 6 (each a different one); our questionnaire's number was 6. A user was assigned to one of the questionnaires 1 to 6 right after opening the link to the survey. Which of the six questionnaires actually got displayed to a user was then defined by that user's player ID: The system assigned

7 A user was considered active if he or she had signed in to the portal at least once in the last year; the survey was only accessible when signed in.

user 1, that is, the player with player ID 1, to questionnaire 1, user 2 to questionnaire 2, until questionnaire 5 had been assigned, after which questionnaire 1 was assigned again, and so on. However, the player ID of a respondent user was not only recorded to calculate this algorithm, but also to collect performance data via web crawler and subsequently match each questionnaire entry with a corresponding performance score (see below).

To maximize the number of responses, respondents were allowed to answer more than one questionnaire (but, of course, never the same twice). As an additional incentive for it, filling out more questionnaires would improve the survey candidate's chances of winning a prize. To ensure that a candidate did not answer the same questionnaire twice, cookies were used.

3.4.1.2 *Collection of Performance Data via Web Crawler*

A user's player level–the measure of performance in this study–can be found on the user's personal profile web page. In order to retrieve the player level of every user who had filled out to our questionnaire as well as to store this information, we programmed our own customized Java-based[8] web crawler. To this end, we first screened the survey data for all recorded player IDs and passed them to the crawler. On the basis of the player ID the crawler then automatically built the individual uniform resource locator (URL) for the profile page of the respective respondent and subsequently downloaded the full content of the corresponding profile page. We configured our download patterns in such a way that they appeared human-like, thereby avoiding triggering the intrusion detection of the portal server while crawling; otherwise, access to the profiles could possibly have been blocked, a fact that would have forced us to switch to another IP. Because the parsing of the profile page data was only done after the crawling instead of simultaneously, we were able to fully concentrate on the crawling task during these activities.

We started the crawling process at the end of April 2012; in total, it lasted nine days. Later we parsed the downloaded files offline, thereby extracting and storing a player level for each user with the help of a parser we had specifically built for this purpose. The whole parsing process was executed in fully automated fashion which–thanks to advanced testing routines–yet still allowed for total control of possible error sources. Also, this approach enabled for customized preprocessing and preliminary data investigation (see Section 4.1 for more details) currently not supported by software tools or programs like SPSS (Statistics),[9] (SPSS) AMOS,[9] or SmartPLS,[10] for example.

8 http://www.oracle.com/technetwork/java/index.html
9 http://www.ibm.com/software/analytics/spss/
10 Ringle, C.M.; Wende, S.; Will, S.: SmartPLS 2.0 (M3) Beta, Hamburg 2005, http://www.smartpls.de/

3.4.2 *Exclusion of Suspended Participants*

Some preparatory actions were necessary before being able to merge our data: While parsing the user profiles stored locally on our hard drives, we discovered that some profiles had had the status “suspended” at the time of the web crawling. Furthermore, some downloaded files were incomplete or even empty. Consequently, it was impossible to retrieve performance data for both of these groups of profiles, and we eliminated the corresponding 163 survey entries. At the end of this step, 6,616 cases were left.

3.4.3 *Merging the Data Sources*

To analyze the impact of the hypothesized predictors of performance simultaneously, data sources of two different origins, resulting from (a) the survey as well as (b) our web crawling activities, needed to be merged into one file. As mentioned before, we matched the survey entry of a user with the according performance information using the user’s player ID as matching key. Again we programmed a tool which executed the matching and merging automatically. As a result, we obtained a file in which each line represented an answer to our questionnaire, with one new column containing the corresponding player level score of the respondent and another new column containing the crawl date. This file was constructed in such a way that it was compatible with the statistics software which we intended to use for data analysis, for instance SPSS and AMOS.

3.5 OUTLINE OF ANALYSIS ACTIVITIES

For a concise overview of our activities performed during the pilot studies and the actual study as well as and their outcomes, see Table 7. The envisaged approach to subsequently analyze our data was to apply SEM. On the basis of our extensive literature review on this method of analysis, we draw the conclusions that

- determining a suitable estimation method should ideally take into account the stage of research (exploratory, confirmatory, replication, etc.), the data distributions observed, and the available sample size, and that furthermore,
- selecting a treatment for the missing values should ideally be dictated by the assumed missing mechanism and other missing characteristics, the variable distributions present in the data, and the analysis procedure to be applied subsequently.

From these conclusions, we inferred intermediate objectives to focus on while pursuing the fulfillment of our overall research goals.

Table 7: Scale development activities and deliverables. Adapted from Churchill (1979) and Limayem, Hirt, and Cheung (2008)

No.	Activities	Deliverables
1	-Find suitable existing scales for primary and secondary constructs, adjust to context; develop construct items and controls.	-List of construct and control items
	-Create item list for gaming preferences, awareness of cheating, motives for playing, improvement suggestions, and demography.	-List of additional candidate items
	-Translate candidate items into German (if applicable); assess content validity with help of operator staff.	-List of items in German for first pilot
2	-Pilot 1: use 155 own and 15 items shared with operator.	-69 cases collected from moderator forum, approx. 50% incomplete;[a] 2 adapted lists of items for next pilots
	-Pilot 2: use 70 own items, solely covering the CIRME construct.	-180 cases collected from community, 90 incomplete[a]
	-Pilot 3: use 187 own and 14 items shared with operator.	-590 cases collected from community, 118 incomplete[a]
3	-Assess quality of instruments based on PCA and item-scale statistics (e. g., Cronbach's α); reformulate/eliminate inadequate items.	-Reduced list with 103 construct items
	-Create new items to control for consistency of CIRME and new preference item.	-5 new controls (incl. 1 demographic item), 1 new preference item
4	-Final survey: use 126 own and 15 items shared with operator.	-6,779 cases collected from community, 3,417 incomplete[a]
	-Web crawling: gather performance information from profiles, document crawl date, clean data.	-6,616 data points for performance collected
	-Merge all data sources (survey data, performance information, and crawl date) into one file.	-6,616 merged cases in total

[a] As reported by the survey tool.

This chapter has first described the context of the study and presented details about the its participants. Second, it has dealt with the development of measures, the measurement model, and the data collection process. Finally, it has outlined the upcoming steps of analysis and how they relate to each other. The next chapter presents the data preparation, validation processes, and hypothesis testing.

4 DATA SCREENING, ANALYSIS, AND RESULTS

There is nothing more deceptive than an obvious fact.

— *Arthur Conan Doyle*

The previous chapter has described how the study was conducted, how measurements were operationalized, and how data was collected. In this chapter, we first illustrate the examination and preparation of the final data and examine the psychometric properties of our measures. We then proceed to explaining the process of hypothesis testing and presenting our results.

4.1 PRELIMINARY EXAMINATION AND OMISSION OF DATA

To comply with good scientific practice, our analysis started with detecting and diagnosing data abnormalities and faulty data (cf. Malone & Lubansky, 2012; van den Broeck, Argeseanu Cunningham, Eeckels, & Herbst, 2005). Examination showed that several entries in our data file had the same value for the player ID field, so that some users were associated with more than one set of answers in our data base. As the player ID is designed as a key and thus a unique identifier, this meant that some users had answered our questionnaire more than once. Our observation led to the conclusion that in some cases, the monitoring via cookies had been levered out–either because participants did not use the same computer every time they had logged onto the portal, or because they had deleted their cookies in between two questionnaire rounds. In total, we counted 6,058 distinct player IDs, out of which 493 were found more than once in the data. The latter group of participants–that is, those with "duplicate" IDs referred to as the nonunique group hereinafter–accounted for 1,051 cases in the survey file. As $493 * 2 = 986 < 1,051$, we concluded that some IDs should appear even more than twice in the data. Indeed, we found a single player ID up to six times in the file.

Because the situation described above had eventuated as a consequence of technical problems and not of a decision deliberately made by the concerned participants, we opted against a general and unevaluated omission of all cases related to the nonunique group; a complete removal could have lead to an incalculable selection bias and ultimately to an unusual group with regard to the population as a whole (cf. Schafer & Graham, 2002). Yet, irrespective of the fact that including as many entries as possible was a desirable objective in terms of selection bias, applying strong restrictions for case selection from this special subsample seemed very necessary to ensure data quality. An obvious requirement was that at the end, no more than one entry per user was to be retained. To decide which cases to classify as problematic and which ones to finally omit from further analysis (cf. Bagozzi & Yi, 2012), we analyzed the nonunique group separately from the unique group first, in order to develop well-defined selection and elimination strategies before ultimately implementing according procedures. We continue explaining our approach in more detail in the following sections.

To appraise the data quality of this subsample, we first investigated its self-disclosure items (see Section 3.2.2.3). It quickly became apparent that a selection based on self-disclosure was not feasible, because this information was missing for the majority of entries (as shown in Table 8), either due to nonresponse on the self-disclosure items or due to dropout at some earlier stage. We therefore sought for further information on the nonunique group through the analysis of its player levels. The player levels were informative inasmuch as one can generally expect a user with a player level of zero to be a beginner and not much involved in eSports or the community, whereas a user with a high player level is able to boast certain achievements and is likely to exhibit some sort of interest in eSports as such.[1] We therefore presumed that the distribution of the player level variable provided indication as to whether the segmentation into the unique and the nonunique group originated in some sort of self-selection. Moreover, unlike in the case of both self-disclosure items, this variable supplied a data point for every single ID. Because it had been extracted from the respondents profile pages automatically it exhibited no missings at all; as a result, the player level was accessible for evaluation for every study participant. Inspection showed that in the nonunique group, player levels ranged from zero to 294 (see Table 9); they also showed a relatively even distribution and resembled the whole sample in terms of range and distribution, too (cf. Table 12 later in this chapter). These observations gave rise to the assumption that the nonunique group represented a wide range of the population (and not an unusual subgroup like, e. g., only beginners), so that this subsample was unlikely to have formed by self-selection (cf. Heckman, 1979). Another relevant aspect of data quality is the presence and particularly the extent of missing values in the data (cf. Malone & Lubansky, 2012). Even for very small percentages of missings it is discouraged not to use proper MDTs (Graham, 2009). Among the nonunique group, we focused our attention on missings on the primary construct variables. Missing values analysis thereby included so-called system missings–which, in our case, resulted from participants quitting the survey for good–as well as missings which resulted from users ex-

1 At this point it needs to be stressed that, although zero is the lowest possible player level, this score does not necessarily imply that a user never won a game or achieved nothing so far; as described before, the level reflects the achievement of a user in comparison with every user who plays the same game(s), and it takes considerable effort to obtain a level of even one, for example.

Table 8: Missings on self-disclosure items in nonunique group (N = 1,051)

		answers_ caref[a]	survey_ length[b]
N	Valid	104	255
	Missing	947	796

[a] Self-evaluation of the extent to which a participant had honestly answered the questions.
[b] Indicated a participant's rating of the length of the survey.

plicitly choosing "no answer" for a specific item. We then evaluated the effect of being asked to fill out the exact same questionnaire *again* on a participant's motivation to respond. We postulated that, for each additional round a participant was requested to answer our questionnaire instead of one of the other five, the willingness to answer the questions should decrease steadily and be liable for more and more missings each round. We thus accounted for the timestamp of every entry and included their chronological order in our examination. But different than expected, there was no link between the amount of missings and the number of answer rounds a participant had undertaken so far, so consequently, analyzing the exact sequencing of entries of a particular participant did not qualify as a means to advance our investigation of the nonunique group.

Based on our literature review which showed that most data missing techniques have not been tested for more than 50% of missing values,[2] we suspected that keeping cases with a higher percentage of missings than that should not lead to more accuracy, independent of the MDT we attempted to use. Hence we considered 50% of missings on the primary construct variables as a reasonable threshold for selection. Consequently, IDs which did not provide at least one case with 50% of missings at most (or, respectively, a minimum of 50% of valid data) were rejected so that 245 out of initially 493 IDs remained; some of the selected IDs were associated with only one case below the 50% threshold, while other IDs were associated with more than one case below this threshold. Both groups, that is, the single-50%-entry group (consisting of 190 IDs) and the multiple-50%-entries group (consisting of 55 IDs), respectively, were examined further.

The goal of our next step was to exclude users who had answered our questionnaire in an inconsistent manner; more precisely, we aimed at detecting internally inconsistent answers of a participant within one questionnaire round, that is, *local* inconsistency, but more importantly,

2 See Section 5.3 and, for example, McKnight et al. (2007), on the ambiguous use of the term amount with reference to missings in a data set, that is, amount per variable vs. per case vs. over all.

Table 9: Player levels for all distinct IDs in the nonunique group before final selection ($N = 493$)

Level	n	Valid %
0	20	4.1
1–10	87	17.6
11–20	86	17.4
21–30	77	15.6
31–40	64	13.0
41–50	44	8.9
51–60	34	6.9
61–70	28	5.7
71–80	14	2.8
81–90	14	2.8
91–100	5	1.0
104–294	20	4.1

Note. This particular group of IDs accounted for 1,051 entries in the nonunique data file in total.

also across different answer rounds, that is, *global* inconsistency. In the single-50%-entry group, the one case that met the threshold was compared with the respective cases that had exceeded the threshold; for the multiple-50%-entries group, the "best" case (the one with the least missings on the construct variables) was compared with its corresponding cases that did not meet the threshold. Although the approach was similar, the groups were examined separately in order to obtain a more detailed view of the response behavior and the resulting data quality.

With the aid of several distinct variables which were relatively simple to compare, that is, age and gaming preferences, we started with examining local consistency; the goal of this analysis was to reveal whether age statements or answers on gaming preferences within a particular entry were consistent or appeared to be contradictory. We then checked the same variables across all entries that belonged to the same ID. The reasoning behind this approach regarding age was that a maximum of +1 in age difference was justifiable, as the user could have had his or her birthday between two rounds (but only once). In such a case, the timestamps needed to support this assumption, and a user could not have gotten "younger" afterwards again, for example. Accountable +1 cases related thereto were only found twice in the analysis. Relating to gaming preferences, our reasoning was that, due to the effort one has put into a game to be good at it, and so on, due to the effort one has put into a game to be good at it and so on, it is unlikely for a user to like a particular game best at a certain time and to prefer a totally different game after only a few days or weeks. Only if all answers were fully consistent, the ID under investigation was kept. Out of the 245 IDs at the beginning of this procedure, 214 IDs remained at the end. To check for global consistency, we looked for discrepancies between answers of respondents across different survey rounds, focusing on the primary construct variables. In this vein, we compared the–possibly deferring–rating scores which a participant had given for a particular item across rounds; as for instance, a participant could have given a rating of "strongly disagree (-3)" for a specific CIRME item in the first round and a "strongly agree (+3)" for the same item in the second. We performed such a comparison with every primary construct variable and every possible combination of rounds: The rating score of the first round was compared with the rating score of the second round, the second compared with the third, the third compared with the first, and so on. For every ID in the nonunique group and for every variable, we then computed the distances of the ratings across rounds and ranked the overall degree of rating deviance using an evaluation system we had developed for that purpose. An ID was then kept or removed on the basis of its total deviance score. As mentioned before, for every ID that was kept at the end of this process, the entry with the least amount of missings was finally selected to be analyzed. In the end, 178 IDs–and consequently the same amount of cases–were selected from the nonunique group for further analysis. Their data was then merged with the data of the 5,565 unique IDs, so that in total, the cases in the resulting data file summed up to 5,743. Table 10 summarizes the selection process.

4.2 SPECIFICS OF PARTICIPANTS AND DATA

This section first describes the main demographic characteristics of the merged sample and analyzes extreme values of these characteristics (i. e., outliers), in order to provide a better understanding of the sample's nature. It then proceeds to presenting more universal details such as data distributions and certain aspects of missing values (e. g., extent and patterns) related to our data.

4.2.1 *Demographic Characteristics*

We expected awareness of privacy issues in this particular target group–probably best described as "technophiles" undergoing the constant debate on privacy in Germany–to be very high and suspected that only few information on demographic sample characteristics would be ob-

Table 10: Selection of entries from nonunique group throughout the different selection steps

Steps		Initial # IDs	Description	Remain. # IDs
1	a	493	Exclude IDs with more than 50% missings	245
	b		Split into a single-50%-entry group and a multiple-50%-entries group	
2	a	190	Check single-50%-entry group for inconsistency through analysis of age and other variables	165
	b	55	Same check for multiple-50%-entries group	49
3		214	Check for inconsistent answers by comparing construct variables across all survey rounds through rating scores	178

tained, even though participants were assured of their anonymity. The evaluation of our data indeed showed a considerable amount of missing values on the respective variables, with the exception of age, as we explain below. Table 11 shows the degree of missings on selected demographic variables of interest to the study. The following figures refer to *valid* percent, meaning that they only account for participants who actually answered to corresponding questions, so that they only inform about percentages with reference to that group and thereby ignore nonrespondents. Out of the former, 95.9% indicated to be male, the other 4.1% to be female; 56.5% of the participants were single, while 37.6% were in a relationship but unmarried, 4.1% married and again

Table 11: Missings on selected demographic variables (N = 5,743)

		Age[a]	Age[b]	Gender	Personal status	Occupation
		age	*dem_1*	*dem_2*	*dem_3*	*dem_4*
N	Valid	5,513	978	962	916	910
	Missing	230	4,765	4,781	4,827	4,833

[a] The variable *age* used predetermined clickable age categories from "6 years old" to "41 years old or older".
[b] The variable *dem_1* was a free text field.

1.7% were either divorced or widowed; 32.1% of the participants went to school, 25.8% pursued an apprenticeship or received professional training, 25.7% were employed or self-employed, 10.6% were students, whereas the remaining 5.8% either fulfilled a civil, military or voluntary service, were unemployed or had taken a childcare leave (totals of percentages are not always 100 because of rounding). Among others, we present figures displaying on the participants' age in Table 12.

4.2.2 *Univariate Outlier Inspection*

According to Yuan and Bentler (2001), even a small proportion of outliers may pose a problem when using classical statistic procedures. In our study, because most of our items were measured via rating scales and thereby bound, the only variables sensitive to univariate outliers and extreme values were the variable measuring the player level as well as the two age variables. In the following, we first explain the existence of two age variables in the survey.

While the operator's variable *dem_1* captured age via a free text field and was optional, our *age* item was instead sampled using predetermined age categories (from "6 years" to "41 years or older"), to be chosen from via mouse click. Ours was also a forced item, to avoid writing or other errors as well as "no answer" missings. With the exception of extreme values (due to nonsense or unlikely answers of the age variable *dem_1*) together with discrepancies related to differing start and end values (e. g., the first category of our age variable *age* was "6" and the last category "41≤"), the distributions of *dem_1* and *age* were almost exactly the same when comparing valid and cumulative percentages for both. For this reason, we focused on our *age* variable for analysis and took *dem_1* only into consideration for the purpose of cross-validation.

For consistency with regard to our hypothesis and the postulated relationships between the constructs, we excluded all participants which were younger than 14. In the end, after removing the participants who had indicated nonsense or unlikely age values as well as those who were too young for the study, 5,588 cases of originally 5,743 remained (see above). Age and player level distributions of the remaining data sample are depicted in Table 12. As can be taken from there, player levels ranged from zero to 482. The player level variable, as mentioned previously, exhibited no missings at all owing to the method applied for its data collection (i. e., via web crawler). Because our research goals did not imply restrictions concerning player levels–in contrast to restrictions concerning age–an exclusion of participants on the basis of this variable was not useful at this point. Also, literature suggests that outliers can be detected from the examination of the univariate distributions (e. g., McDonald & Ho, 2002). As we therefore needed to generate information on the player level distribution first, the analysis

Table 12: Age and player levels for the sample population after elimination of age outliers (N = 5,588)

Characteristic		n	Valid %
Age[a]	14–16	791	14.7
	17–20	2,316	43.2
	21–25	1,620	30.2
	26–30	424	7.9
	31–35	126	2.3
	36–40	42	0.8
	41≤	45	0.8
Player level	0	348	6.2
	1–10	1,171	21.0
	11–20	1,089	19.5
	21–30	901	16.1
	31–40	616	11.0
	41–50	414	7.4
	51–60	320	5.7
	61–70	221	4.0
	71–80	131	2.3
	81–90	108	1.9
	91–100	71	1.3
	101–482	198	3.5

Note. Totals of percentages are not 100 because of rounding.
[a] Totals of frequencies for this variable are 5,364 due to 224 missings.

of possible outliers on this variable and, related therewith, the analysis of multivariate outliers was postponed. In the upcoming section, we present the results of the distribution examination. We performed the latter not only for the purpose of identifying outliers and evaluating our options with regard to subsequent methods of analysis, but also to evaluate our options with regard to MDTs, as explained in the following.

4.2.3 *Distributions and Missing Data*

The system reports and our above-mentioned data screening activities had already drawn our attention to a considerable amount of missing values in our data. To avoid biased estimates, the treatment of missings had to *precede* the examination of item reliability or psychometric properties (Bühner, 2011; Enders, 2003). Following our literature review, the next steps consisted of investigating the measure distributions of the variables of interest first (in our case, those of our constructs) and then proceeding to analyzing important characteristics of their missing values.

Our reasoning on the adequacy of available MDTs for our particular data on the basis of our missing values examination and regarding the procedures we intended to apply (e. g., for the assessment of the psychometric properties of our measurement and model testing) is presented below. However, we further elaborate on the subjects treated here later in the DISCUSSION chapter.

DISTRIBUTION EXAMINATION Under conditions where normality distributions are violated, maximum likelihood (ML) standard errors and test statistics based on normal theory are liable to bias (Yuan, Bentler, & Zhang, 2005). Authors have yet frequently highlighted that most data sampled by means, for instance, of rating scales will fail to pass strict tests of normality like the Kolmogorov-Smirnov test and the Shapiro-Wilk test (see Weiber & Mühlhaus, 2010, p. 147). In particular, normality tests are not useful for items with discrete answer formats (Bühner, 2011, pp. 233-234) and when dealing with a large sample (Bühner, 2011; Weiber & Mühlhaus, 2010, see also Section 5.3). It was therefore not surprising that our construct variables all failed to pass the aforementioned normality tests (significance level: $p < .001$).

Tests for multivariate normality like Mardia's coefficient are similarly sensitive to very small deviations from normality, especially when–again–samples are large (cf. Kline, 2011). Testing our data for multivariate normality became even more of an issue due to missing values, because tests which simultaneously account for missing data are not yet implemented in software like AMOS , for example (IBM AMOS Support, 2009; see Yuan, Lambert, & Fouladi, 2004, for the prospective development of multivariate normality test in the presence of missing data). For the purpose of checking for departures from univariate and multivariate normality, literature therefore suggests to rely on the analysis of skewness and kurtosis values (e. g., Weiber & Mühlhaus, 2010; Kline, 2011; West, Finch, & Curran, 1995; Bühner, 2011) as well as on visual clues provided by graphical representations of variable distributions (Bühner, 2011; see also Kline, 2011, p. 60). Tables displaying skewness and kurtosis values for our primary and secondary construct variables as well as the histograms and normal Quantile-Quantile (Q-Q) plots for the primary construct variables can be found in Appendix B, Section B.2 (on how to read them, see also e. g., Kline, 2011, p. 209; and Wilk & Gnanadesikan, 1968, p. 5).[3] Except for four of them, kurtosis and skewness values were in the range of $|{<}1|$ for all construct variables.

3 Due to lack of space, only tables but no diagrams are presented for the secondary constructs.

Out of the aforementioned four variables, three only just exceeded this very strict threshold with regard to kurtosis (i. e., *internwo*10_14_2: -1.000, *internwo*10_14_3: -1.079, and *rules_parents*_2: -1.021) but were still below the threshold in terms of skewness, and only one variable, namely the performance indicator *esllevel* (skewness 3.323, kurtosis 20.557), was found to considerably exceed even more relaxed threshold values for both parameters published as in literature (i. e., |<2| for skewness and |<7| for kurtosis, see Weiber & Mühlhaus, 2010, p. 146; or, alternatively, |<3| for skewness and, as a "conservative rule of thumb", |<10| or respectively less conservative, |<20|, for kurtosis, see Kline, 2011, p. 63).

Thus with the exception of the performance measure, no serious departure from normality of any construct variable was found, and because the majority of our variables had univariate skewnesses and kurtoses in the range of -1.0 and +1.0, "not much distortion" (B. O. Muthén & Kaplan, 1985, p. 187) was to be expected (i. e., even if *not* using proper MDT or of *not* using analysis methods with relaxed distribution assumptions, q. v.).

MISSING VALUES ANALYSIS Next, we performed an analysis of missing values.[4] Hereafter, missings refer to all types of missing values related to our data collection (see, e. g., McKnight et al., 2007; Schafer & Graham, 2002; and Andridge & Little, 2010), that is, from missings for individual items (also called item nonresponse) to missings originating from dropout at any stage of the questionnaire, either due to a conscious decision of the user (i. e., quitting due to survey fatigue, boredom, etc.) or resulting from negligence, technical problems, and so on (van Ginkel, van der Ark, & Sijtsma, 2007; Allison, 2003; van den Broeck et al., 2005, give examples of reasons for missings).

Simply excluding cases from analyses for which available data are not complete, thus performing a so-called complete case analysis, may introduce such systematic bias that researchers potentially draw wrong conclusions from the results of a study (e. g., Schafer & Graham, 2002; Nakagawa & Freckleton, 2008). But despite the fact that missing values are such a common problem, there still seems to be no simple, straightforward, or comprehensive approach for choosing an appropriate treatment of the missings in a specific data set. From what we were able to discern from literature, to this end, aspects which should affect the decision on what MDT to apply are

1. the hypothesized so-called missing mechanism, an aspect considered essential;

2. the data distributions, the amount, the patterns, and the possible types of missings;

4 Reminder: As we doubted the value of their responses, we deleted 11 age outliers containing nonsense values, so that these cases were excluded before analyzing missings.

3. the method of analysis to be applied to the data, which is usually determined by the type of research questions a researcher seeks to answer;

4. and finally, the interaction of the three aspects above.

According to the widely known categorization system first introduced by D. B. Rubin, a very strict assumption called missing completely at random (MCAR) holds if, roughly speaking, the fact that a particular item is missing for a particular participant can be neglected because the "true" value of that missing item is not related to any other characteristic of that participant. To make a very simplifying example, consider an item capturing income that is the only item with missings of a particular questionnaire; if it were missing for participants with *very low, low, moderate, high,* and *very high* income on an equal share with reference to the whole sample and with no correlation to answers given to other items in the questionnaire, one can probably assume MCAR.[5] A missing mechanism with a more relaxed underlying assumption in terms of statistical randomness is missing at random (MAR). in turn, missing not at random (MNAR) does not assume randomness and is generally considered "nonignorable"–as opposed to MCAR and MAR–because the missingness represents relevant information about the sample (e. g., Allison, 2003; Grittner et al., 2011; van Ginkel et al., 2007).

Throughout literature, the missing mechanism underlying a set of data receives a lot of attention because of its hypothesized impact on the performance of MDTs. Findings suggest that under typical data conditions, the use of listwise deletion, pairwise deletion, mean and hot deck imputation, regression substitution–a group of MDTs often subsumed under the term of ad hoc procedures (see Tsikriktsis, 2005, for an extensive list of these procedures) as well as the use of the so-called missing-indicator or missingness dummy variable method, respectively, will typically lead to unsatisfactory results (Collins et al., 2001; Enders, 2001a, 2001b; Kamakura & Wedel, 2000; Savalei & Bentler, 2009; Sinharay et al., 2001; Donders, van der Heijden, Stijnen, & Moons, 2006; Graham, 2009; Allison, 2009) and to lower power and validity of tests (Allison, 2003; Roth, 1994; for differing propositions under specific circumstances, see Graham, 2012a; Roth, 1994; Mackelprang, 1970). For example, if at all, listwise deletion is only considered "safe" to use under the MCAR condition (Schafer & Graham, 2002; Sinharay et al., 2001; Baraldi & Enders, 2010); similar restrictions apply to the aforementioned MDTs. However, the MCAR assumption is yet very unlikely to hold with real-life data (Allison, 2003; Enders, 2001b; Myers, 2011).

Many modern MDTs with properties advantageous to those of ad hoc procedures assume MAR (cf. e. g., Collins et al., 2001; Enders & Banda-

5 The violation of the MCAR assumption can be tested: Little (1988) has proposed a multivariate χ^2-test for MCAR, which is also implemented in SPSS; it is, however, sensitive to departures from normality.

los, 2001; Sinharay et al., 2001; Allison, 2003), so a less restricted missing situation than MCAR. However, "there is no way to test whether the MAR assumption is violated" (Allison, 2003, p. 555), because that would paradoxically require the assessment of the true values of the missings, thus exactly of those observations which are not available (McKnight et al., 2007; Raykov, 2012). On the other hand, although differing opinions have been expressed (e. g., Roth, Switzer, & Switzer, 1999), it has been proposed that studies in the social sciences are less likely to be liable to MNAR–in contrast to clinical studies where dropout is often "informative" (Goyal & Gomeni, 2013, p. 1570) because "participants may be dropping out for reasons closely related to the outcomes being measured" (Schafer & Graham, 2002, pp. 172–173). In fact, some even argue that dropout due to boredom at the end of a trial can be considered MCAR (even in clinical research, cf. Zhang, 2005). And finally, though departures from MAR should be suspected for many cases, a MNAR situation does not necessarily imply that estimates and standard errors will seriously be affected when using MAR-based MDTs (Schafer & Graham, 2002). For more information, the reader is referred to the upcoming chapter.

The characteristics of the missings in our data can be summarized as follows: When testing our construct variables in their entirety, we found (and, given the likelihood of MCAR as discussed above, were not surprised to find) that their data violated MCAR (significance level: $p < .001$).[6] With reference to *amount*, percentages of missing values on single item variables in our data ranged from 36.7% to 54.2%. With reference to *patterns*, we found that by analyzing patterns for all variables simultaneously, most patterns (86.5%) were exclusively monotone (see Schafer & Graham, 2002), that is, when items were analyzed in the same order as they appeared in the questionnaire;[7] these findings together also indicated that–with reference to *types*–the great majority of missings occurred due to dropout of participants who quit the survey at different stages of the questionnaire. Tables displaying the amount of missings and visualizations of their patterns are presented in Appendix C, in Section C.1 and Section C.2.

In the upcoming sections, we go into further detail on how we determined which MDT to apply to our data, with special regard to the analysis procedures we envisaged to use subsequent to that.

6 For reasons of completeness: Some variable subsets of the secondary constructs actually passed the MCAR test, namely those of the parental control construct as well as those of the inter-reality mediation construct.

7 Patterns with less than 1% cases (i. e., 56 cases or less) were discarded to reduce complexity.

4.3 MISSING DATA TREATMENT FOR EXPLORATORY ANALYSIS

So far, we have presented the results of the distribution inspection for our construct variables, our reflections on the likelihood of certain missing mechanism underlying our data as well on amount, patterns, and types of our missing values (i. e., of our construct variables). As with previous topics, we present more details relevant to the context of the upcoming sections later in the DISCUSSION chapter.

In order to identify a treatment adequate for our missings it was necessary to figure out which requirements and assumptions of available high-performance MDTs would match with our data and missing characteristics, thereby keeping in mind the analysis procedures we had envisaged. Considering the characteristics of our data,

- the use of ad hoc procedures seemed inadvisable per se, because–not surprisingly–the MCAR assumption did not hold with our data; according to our investigation, however, it seemed appropriate to consider MDTs which assume MAR; also,
- variables which exhibited missings were found to follow a normal distribution within the recommended limits, while the only variable that exceeded these thresholds was also the only dependent variable and did not exhibit any missings (cf. Sterne et al., 2009, and below); therefore, the application of MDTs which assume a normal model seemed appropriate; and finally,
- studies have confirmed that many MDTs perform very well even at the 50% level (e. g., Kamakura & Wedel, 2000); in our case, the maximal amount of missings on a single variable only exceeded 50% by very little.[8]

With regard to the analysis procedures to be applied, as the next step was to perform a reliability analysis in order to reexamine our measures, we needed to find an MDT suitable for this particular application first.

Two MDT approaches are considered state-of-the-art today, namely procedures using multiple imputation (MI) or, alternatively, procedures based on ML, the latter of which includes full information maximum likelihood (FIML) computations and procedures which make use of the expectation-maximization (EM) algorithm (cf. Schafer & Graham, 2002).[9] Both approaches produce "very accurate results" under the MAR assumption and on the grounds of a formal probability model

8 As we explain below, we additionally deleted cases which had no data on the variables of interest during the course of the analysis; this reduced the maximal amount of missings for a single variable to 39.39%.

9 Note that this widely cited paper specifically refers to MI as "Bayesian multiple imputation" (p. 147), as opposed to other approaches based on different techniques; see also Section 5.3.

like the normal model (Sinharay et al., 2001, p. 318). With these modern MDTs, single missing items can be treated *before* forming complete scales (Schafer & Graham, 2002; Roth et al., 1999; Enders, 2003)–the preferable approach (Graham, Horn, & Taylor, 2012) when contrasted with generating latent variable (or scale) scores on the basis of all the available data on the observed variables (Campbell, Rijsdijk, & Sham, 2007; cf. also Graham, 2009).

The next steps planned consisted of conducting an analysis on the basis of Cronbach's α and associated statistics and performing an exploratory factor analysis (EFA); for this kind of analysis, the use of EM-based MDT has been recommended repeatedly (e. g., Enders, 2003; Graham, 2009), hence we intended to follow this recommendation. EM-based MDT generate a full data matrix by imputing values in a specific manner. But because the EM algorithm implemented by SPSS does not correct the imputed values when treating missings, thus leading to a certain bias when applied (Graham, 2009), we first needed to search for another program capable of performing said corrections. We found it in the NORM program,[10] the latter of which provides the required data augmentation by simulating random values of parameters and adding them to the values recently imputed.[11]

Another recommendation with regard to treating missings is to remove cases that have no data on the variables of interest from the imputation procedure (e. g., in the context of MI, see Graham, 2012e). We therefore separated all cases of our data into two groups such that one group contained the cases which had data on at least one value of a primary construct variables and the other contained the cases which had only missing values on these variables(cf. Appendix C, Section C.2), in order to control for a selection bias by means of comparing the two subsets with regard to the groups' age and player level distributions; the results of this comparison can be found in Table 13 (Sample 1 contains data on at least one primary construct variable, whereas Sample 2 contains no data on these variables at all). As one can see from there, the distributions of both variables were quite similar across both groups. An interesting observation can also be made when looking at the player level distribution, as percentage for level zero is slightly higher for those cases with no data than for those with data; in our view, this supported our earlier argument that there possibly exists a link between users' player levels and the level of interest they take in community activities; those who have achieved a certain player level seem to be more likely to engage more in community activities like, in the case at hand, a community survey project. These findings encouraged us to continue

10 NORM: Multiple imputation of incomplete multivariate data under a normal model (Version 2) [Software] (1999). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu.

11 Schafer, J.L. (1999) NORM users' guide (Version 2), section "How can I create imputation for exploratory purposes?." University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu.

on the current path and to exclude the cases that exhibited no data on the primary construct variables from further analysis.

We then exported our SPSS data file with the remaining 4,219 entries to the file format required by NORM, read the data into the program, and followed the recommended procedure to obtain EM estimates (Graham, 2012e). In this vein, we not only added all 67 primary construct variables including the player level to the EM imputation model, but also the variables of the secondary constructs, the controls, and so on (except for those with an unsuitable scale level[12]), all summing up to 118 in total. This way, on the one hand side, the imputation model–which serves to reflect the information on relationships between missings available from the data as a whole–accounted for all variables that may correlate with the variables of the actual research model (cf. e. g., Graham, 2012e); by including additional, so-called auxiliary variables, a strategy which is often termed *inclusive* (Collins et al., 2001), we aimed at making the MAR assumption more plausible and improving the performance of EM (Enders, 2003). On the other hand side, all dependent (or outcome) variables, in our case the player level, were taken into account for the imputation a means to address the fact that missings are not totally independent of each other (i. e., that they are not MCAR, cf. Sterne et al., 2009). At the same time, because data on this variable was available for every participant, imputing values for a dependent variable–a practice which has been subject to discussion in literature (Graham, 2003; Allison, 2009), was not needed. EM converged normally in 116 iterations (for comparison, running the imputation with only the primary construct variables needed 28 iterations). Supported by the NORM program, all imputed values were then corrected through some random error terms, an important step recommended to reduce bias (Graham, 2009). The NORM reports covering the updated percentages of missings and the most frequent patterns remaining after the elimination of cases with no data on the primary construct variables can be found in Appendix C, Section C.3.

MULTIVARIATE NORMALITY AND OUTLIERS ASSESSMENT In theory, because we now had a full data matrix available, assessing multivariate normality and outliers was possible. Unfortunately, we could not expect the results of such an assessment to be trustworthy, for three reasons: (a) Even though its values were adjusted by some random error after imputation, the obtained data matrix only represented a single draw from the sample (as opposed to MI where multiple imputations are performed, see below). The result of an imputation is not deterministic and thus varies after each run, often substantially (Allison, 2012).

12 According to Graham (2012e), the variables are required to be “continuous (or reasonably assumed to be continuous, including ’ordered categorical’ variables), or dichotomous categorical” (p. 71); see also Section 5.3 on this subject.

Table 13: Comparison of age and player levels of cases with data on primary construct variables vs. cases with missing values on all primary construct variables (N = 5,588)

Characteristic		S1 (n = 4,219) n	S1 Valid %	S2 (n = 1,369) n	S2 Valid %
Age[a]	14–16	614	14.6	177	15.3
	17–20	1,843	43.8	473	40.9
	21–25	1,273	30.3	347	30.0
	26–30	329	7.8	95	8.2
	31–35	95	2.3	31	2.7
	36–40	29	0.7	13	1.1
	41≤	24	0.6	21	1.8
Player level	0	246	5.8	102	7.5
	1–10	873	20.7	298	21.8
	11–20	828	19.6	261	19.1
	21–30	695	16.5	206	15.0
	31–40	469	11.1	147	10.7
	41–50	316	7.5	98	7.2
	51–60	249	5.9	71	5.2
	61–70	167	4.0	54	3.9
	71–80	88	2.1	43	3.1
	81–90	77	1.8	31	2.3
	91–100	57	1.4	14	1.0
	101–277	143	3.4	42	3.1
	278–482	11	0.3	2	0.1

Note. Totals of percentages are not 100 because of rounding. S1 = Sample 1 (with data on at least one primary construct indicator); S2 = Sample 2 (with no data on any of the primary construct indicators)
[a] Totals of frequencies are 4,207 (1,157) due to 12 (212) missings.

As a consequence, when calculating indexes of multivariate outliers and normality for an EM-imputed data matrix (like Mahalanobis distance and Mardia's coefficient, cf. Kline, 2011; Yuan et al., 2004; Weiber & Mühlhaus, 2010; West et al., 1995; Peter, 2005), the computation only takes a single random draw from the sample into account–with no way to determine its representativeness with regard to the actual sample values. By the same token, considering that entities under investigation may have missings on many different variables which all have been imputed, eliminating outliers detected via test scores based on these values is equally random. Also, (b) due to the assumption of multivariate normality underlying ML procedures, data will somewhat be "pushed" towards normality when undergoing an ML-based treatment (cf. Bernstein & Teng, 1989, p. 467) such as an EM-based MDT. Consequently, key figures for univariate or multivariate distributions after the use of EM may not be very meaningful. And finally, (c) statistical power of tests like tests of multivariate normality increases with the size of the sample (cf. e. g., Bagozzi & Yi, 2012); for a large sample like ours, a very high sensitivity against departures from multivariate normality was to be expected, which called the validity of such a test for our sample very much into question.

Albeit dubious for the aforementioned reasons, we still investigated Mahalanobis distance, Mardia's coefficient, and critical ratios generated for our EM-imputed data in AMOS. Consistent with the findings of our literature review and our reasoning above (and thus not surprisingly), results showed that our data violated the multivariate normality assumption to a notable degree and contained a considerable amount of multivariate outliers (the player level variable was thereby not included in aforementioned investigation). The following exploratory analysis for the purpose of scale purification and refinement was performed on the basis of the full data matrix retrieved from the EM-based imputation procedure described above.

In preparation of the analysis, the data first had to be transferred back into SPSS-readable format by a program we had specifically developed for this purpose; the latter parsed the output data of the NORM program in order to unify the various delimiters detected (blanks, tabs, etc.), so that the final output data only made use of a single delimiter. This way, the SPSS program was ultimately able to interpret the information in the data correctly.

4.4 EXPLORING FACTOR STRUCTURES AND RELIABILITY

The examination of items and scales is not only recommended for the purpose of purifying the measures before the actual collection of data on the basis of the pretest data, it is also advised to perform this kind of examination with the final data (e. g., Weiber & Mühlhaus, 2010; MacKenzie et al., 2011; Churchill, 1979). To this end, we conducted several EFAs–individually with the data of every construct, but also simultaneously first with all primary constructs and additionally with all constructs together–in order to initially determine factor structures, refine measures, and assess construct validity in a preliminary evaluation (Conway & Huffcutt, 2003; Weiber & Mühlhaus, 2010). Then we assessed reliability on the basis of Cronbach's index of internal consistency first (cf. Cortina, 1993), also known as Cronbach's α. Though we contemplated to exclude problematic items in order to achieve "satisfactory reliability" (Bagozzi & Yi, 2012, p. 14) for the remaining items in an iterative process, we were aiming at attaining meaningful scales and constructs rather than at the maximization of indexes (cf. Bühner, 2011; MacKenzie et al., 2011). For now we only focus on results of the reliability analysis obtained by EFA on the basis of Cronbach's α. When subsequently cross-validating our EFA results via confirmatory factor analysis (CFA), we also assessed validity and reliability of our measurement model on the basis of further indexes, both at the item as well as at the construct level (cf. T. A. Brown & Moore, 2012; MacKenzie et al., 2011; Fornell & Larcker, 1981; Bagozzi & Yi, 2012; Weiber & Mühlhaus, 2010), however, due to the fact that CFA represents a form of hypothesis testing (Backhaus, Erichson, & Weiber, 2011; some note that the distinction between exploratory and confirmatory is not always indisputable, see J. C. Anderson & Gerbing, 1988), we first needed to apply a different type of MDT to our data to be able to do so. Before demonstrating our application of CFA and reporting its results, we explain the MDT applied for hypothesis testing in more detail in the respective section.

For the purpose of exploring the factor structure of our data, we first examined its suitability for factor analysis on the basis of the measure of sampling adequacy (MSA), Bartlett's test of sphericity, and the Kaiser-Meyer-Olkin (KMO) criterion (cf. Bühner, 2011; Bortz, 2005; Patil, Singh, Mishra, & Todd Donavan, 2008; see also Weiber & Mühlhaus, 2010). Next, we needed to determine the number of factors to extract for a construct. Aiming at satisfying all requirements of state-of-the-art publications on the subject, we based our decision on two recommended factor retention criteria (a) parallel analysis (PA) in combination with PCA and (b) Velicer's minimum average partial (MAP) test (see Section 5.3 for details). The amount of factors initially hypothesized for all primary constructs are being compared with the amount of factors we actually found support for by means of PA and MAP test in Table 14; relating to the MAP test, we report results of the original test of 1976 as well as those of the revised test of 2000. Note that the three emotion capability constructs are being first presented separately and then jointly as emotional capabilities, as explained below.

Because (a) oblique rotation is considered superior[13] to orthogonal rotation and (b) as means to allow for factors to correlate (cf. Bühner, 2011),[14] we chose to carry out all extractions with promax rotation. We conducted two factor extractions for each construct (for the emotion

13 Especially superior to the orthogonal varimax rotation, even though the latter is still dominant in EFA practices, (cf. Conway & Huffcutt, 2003); as to oblique rotation as such, promax represents the "method of choice" (p. 338 Bühner, 2011, in original: "Methode der Wahl").

14 This seemed appropriate, as the indicator sets of the same reflective construct are expected to do so (cf. e. g., Weiber & Mühlhaus, 2010). In the process of the analysis this assumption seemed to hold with our data, as all intercorrelations among the factors of our constructs were >.10–with the exception of two factors of the IR emotional capabilities scale, the latter of which exhibited slightly lower correlations

constructs, we also performed a joint analysis), thereby applying two different and particularly recommended extraction methods, namely principal axis factoring (PAF) and ML. Our goal was to compare the resulting solutions and to account for the amount of remaining nonredundant significant residuals (i. e., those with absolute values greater than .05), to ultimately identify the best of two possible solutions in case of doubt (while naturally always bearing interpretability and the initial conceptualizations in mind, cf. Bühner, 2011). In general, ML extraction provided solutions with the least remaining nonredundant residuals, yet in one case, PAF produced an equally good solution, and in another, the ML solution was even slightly inferior to the PAF solution (both with regard to remaining residuals). Eigenvalues, percentages of variance, total variance for all extracted factors of constructs, individual factor loadings of items, and factor intercorrelations obtained on the basis of the *initial* set of items, so before elimination,[15] are displayed in Appendix D, Section D.1. As Bühner (2011) recommends to use ML extraction if an EFA is to be cross-validated via CFA (p. 349), we only present results of the ML extraction and not those of the PAF.

Subsequently, we computed Cronbach's α for each item as well as for each extracted factor (cf. Weiber & Mühlhaus, 2010; MacKenzie et al., 2011), that is, we computed Cronbach's α and so-called item-scale statistics at two different levels of analysis, the subdimension unit-level (cf. Gignac, 2013)[16] as well as at the item unit-level.[17] Results of this reliability analysis, that is, mean-inter-item correlations, Cronbach's α (unstandardized as well as standardized), corrected item-total correlation, squared multiple correlation (SMC), and Cronbach's α "if item deleted" (cf. Bühner, 2011; Weiber & Mühlhaus, 2010; MacKenzie et al., 2011), are presented in Appendix D, Section D.1.2.[18] We subsequently discuss some of the key findings with regard to factor structures and reliability.

Following Joseph and Newman (2010), we had hypothesized three subconstructs of IR emotional capabilities, yet according to PA and MAP, the emotion perception indicators were possibly two-dimensional; for

---

with some of the other factors of the construct, that is, after ML extraction of six factors.

15 Except for the reversely coded items of IR emotional skills, as explained in the following.

16 For a criticism of estimating Cronbach's α at the item unit-level if items are associated with a multidimensional model, see Gignac (2013) and Section 5.3.

17 In this context, Cortina (1993) stresses that Cronbach's α has specifically been developed for estimating reliability "when item-specific variance in a unidimensional test is of interest" (p. 103), likewise the "uniqueness of the item" (q. v.), or "error factors associated with the use of different items"(p. 98); he also urges researchers to interpret Cronbach's α only in the way it has been intended to (i. e., only in the presence of single common factor) and to be aware of its limitations (i. e., of its sensitivity towards high numbers of items and its indeterminate validity in terms of varying item intercorrelations).

18 Note that Cronbach's α will be considerably lower if estimated at the subscale level as opposed to if estimated at the item unit-level (Gignac, 2013).

the emotion understanding indicators, one, two, three dimensions were suggested, and three or four dimensions were suggested for the emotion management indicators. At first appearance, the results seemed to indicate that when dealing with emotion perception and emotion management, we needed to distinguish between capabilities regarding the emotions of the self and capabilities regarding the emotions of others. On the other hand, removing some indicators would have solved the issues related to emotion understanding. However, we noticed undesired MSA values for several indicators; also, factor correlations proposed by factor extractions were sometimes to be very small or even negative, the latter of which were hard to explain. Additionally, the residuals assigned to the solutions were not satisfactory. Because of the shared content and strong conceptual entanglement of the emotion constructs, in our view, inspection of their discriminant validity was crucial. We thus decided to perform an EFA and related reliability analyses of all of their indicators simultaneously before making any further decisions related to the emotion constructs.

All indicators of all emotion constructs together seemed to be described best by six or more factors. Looking at the data more carefully, it turned out that one of these "factors" consisted of all six reversely coded[19] items from three different subscales we had postulated for the IR emotional capabilities construct (i. e., $underst\_o\_6$, $underst\_o\_7$, $feel\_o\_3$, $feel\_o\_6$, $feel\_man\_1$, and $feel\_man\_2$). An affiliation of them was difficult to justify considering the different facets of emotional skills they covered. Said reversed items tended to correlate more with the same item used to measure a different trait than they did with other items of the same trait, suggesting the presence of common method variance.[20] We had to assume that the "clarity" of this factor structure was due to the fact that a considerable amount of participants had not noticed the reverse coding, while others had; as a result, the resulting distributions for these items had become so similar to each other and yet so distinct from the others that enough "support" for an underlying factor was found. After careful considerations and because there was no way to further assess their validity by the means at hand, we decided to exclude the aforementioned items from further analysis.

We then rerun the extraction determination procedures without them and rather found support for five factors. Some of the initial items were identified as to be looked at more thoroughly (or as potential candidates for elimination, cf. MacKenzie et al., 2011), either due to insignificant loadings on any factor, considerable loadings on more than one factor, or reliability issues. It became apparent that the dimensions of perceiving others' emotions and understanding others' emotions were not as clear-cut as expected: $underst\_o\_4$, $underst\_o\_3$

19 For reasons of simplicity and to rule out interpretation errors, they had all been recoded into "positive" items for analysis.

20 This is somewhat ironic, since reversed scales are typically used in an effort to reduce common method variance (F. D. Davis, 1989, p. 327).

as well as *underst_o_5* exhibited almost equal loadings on both extracted factors. The items *underst_o_8*, *feel_man_7*, *perc_s_2*, and *underst_s_1* obtained low reliability scores with regard to SMC, Cronbach's α, and inter-item statistics; *feel_o_1*, *perc_s_3*, and *perc_s_4* showed low reliability, too, but all the worse, in conjunction with low loadings on any of the five factors extracted (cf. Table 14).

The analysis of self-estimated cognitive ability unexpectedly revealed that a two-factor structure may describe the construct more adequately than a single-factor structure. The discovered subdimensions seemed to represent two facets of the construct worth examining more closely: Apparently participants had made a distinction between feedback on their achievements received in the quotidian–or through day-to-day experienced, so to speak–and feedback received by institutions like schools, universities, and so on. Nonetheless, the *clever_5* item could not clearly be assigned to a single factor; the cases of *clever_3* and *clever_1* were similar, though less significant.

Item *cogabs_6* of the focused immersion dimension of the cognitive absorption construct exhibited low loadings on both factors and performed poorly in terms of reliability; it was therefore considered too problematic to keep.[21] The factor structure was yet not affected by the exclusion of this items after performing a follow-up reliability analysis.

Concerning the competitiveness construct, the dimension of personal standards in comparison with others turned out to be problematic, as one of its items, *pers_stand_3*, showed loadings on all factors of a solution with three factors as well as comparatively low reliability scores; the remaining two items of the factor really seemed to represent a dimension of their own, rather than belonging to the other two dimensions. These two remaining items would possibly not provide reliable measurement for a factor. As a consequence, we had to consider to refrain from taking the latter into account and abandon the plan of modeling competitiveness as a higher-order construct (Weiber & Mühlhaus, 2010); this yet included the possibility of loosing an interesting and possibly enlightening facet of the construct (cf. MacKenzie et al., 2011).

Lastly, problems also occurred with *rules_parents_1* of parental control in terms of MSA and reliability. The findings of the individual EFAs are summarized in Table 14.

21 We later realized that the "noise" caused by *cogabs_6* was likely due to negative wording, see IR emotional capabilities on the subject. This resembled the experience the authors of the original cognitive absorption article had made: "Although our initial assessment of these properties in this empirical study is encouraging, the loading for one of the control items was not at the desired level. Interestingly this was the only item that was reverse scaled, suggesting that perhaps the overall convergence of this dimension might be improved by utilizing a third positively as opposed to negatively worded item. . . . Therefore, we encourage others using the CA construct to consider replacing the reverse scaled item in the measurement of the control dimension in order to develop a scale that exhibits greater convergence" (Agarwal & Karahanna, 2000, p. 687).

Table 14: Overview of construct dimensionality / number of constructs, comparing structures as hypothesized, as supported by parallel analysis and Velicer's minimum average partial test (original and revised) of the initial item set, and as finally retained for individual EFAs

Construct(s)	Hyp.*	PA*	MAP* orig./rev.	Ret.*
Perceive IR Emotions	1	2	2/1	1[a]
Understand IR Emotions	1	3	1/2	2[a]
Manage IR Emotions	1	4	3/4	2[a]
IR Emotional Capabilities	3	5[a]	6/5[a]	5[a]
SE Cognitive Ability	1	2	2/1	2
Cognitive Absorption	2	2	1/1	2
CIRME	4	4	4/3	4

* Number of factors (1) Hyp. = hypothesized, (2) PA = supported by parallel analysis, (3) MAP = supported by minimum average partial test (original/revised), (4) Ret. = retained for individual EFA.
[a] After elimination of all reversely coded items of the initial set of items.

Subsequent to the analysis above, we performed an EFA comprising *all* items of all primary constructs, in order to evaluate loadings of indicators on factors when extracted *simultaneously* (cf. Weiber & Mühlhaus, 2010); analysis results can be found in Appendix D, Section D.1.3. These analyses aimed at examining construct and discriminant validity (cf. Weiber & Mühlhaus, 2010). While construct validity refers to the extent to which indicators of a construct measure what they are intended to measure and simultaneously captures the "degree of agreement of indicators hypothesized to measure a construct and the distinction between those indicators and indicators of a different construct(s)"–in contrast to reliability, which only investigates how much a set of measures is in line with a single construct–(Bagozzi & Yi, 2012, p. 18), discriminant validity "assesses the degree to which the scales are differentiable from each other" (Adams et al., 1992, p. 230).

During simultaneous EFAs, we iteratively excluded items which had performed poorly during the individual EFAs and reliability assessments as well as those for which these results were confirmed, which in the end applied to all problematic indicators previously mentioned except for *underst_o_*3, *clever_*3 and *clever_*1. The number of factors to be extracted for this new subset was again determined by PA and MAP. As the total amount of factors for which we had found support during the individual EFAs with regard to the original item set of the primary constructs was 13, we had assumed to find support for just as many factors in the reduced item set; yet the tests rather pointed toward a

12-/10- (MAP) or 11- factor structure (PA), respectively. Initially, we had started with a set of 66 items; at the very start, we removed six reversely code emotions items, and at the end of the exploratory analysis, our set consisted of 49–sufficiently reliable–items for all primary constructs taken together (37 and 34 for all secondary constructs, respectively). Literature stresses to pay attention to the fact that one may be comparing differing sets of items when taking this kind of approach, and that one needs to make sure not to confound the process of assessing model fit by "changing scales (through item deletion) and constructs in an additive fashion" (W. W. Chin & Todd, 1995, p. 238). We therefore deemed necessary to compare the solutions proposed by the individual EFAs and that of the simultaneous EFAs where they imposed different implications (cf. Jones & Fernyhough, 2007) and to further cross-validate all competing and meaningful measurement models derived (see upcoming sections) in order to compare them. The final decision with regard to factors and items were ultimately based on the results of the subsequent CFAs.

Ultimately, we also performed a simultaneous EFA with the remaining indicators of *all* (i. e., primary as well as secondary) constructs, to validate its outcomes with previous findings, too; during its course, *pers_stand_*3 and *rules_parents_*1 were found to be problematic again. For reasons of space, figures of this very large simultaneous EFA are not displayed.

4.5 MISSING DATA TREATMENT FOR HYPOTHESIS TESTING

Note that some of the topics raised here are revisited later in Section 5.3. The next step was to confirm the factor structures found through EFA by the means of CFA. For this purpose, we needed to go back to the last data set (with all its missings) again and apply a different MDT to it, because data imputed by EM–though the first choice for the purpose of conducting an EFA or a reliability analysis–is not recommended for hypothesis testing; standard errors, for example, will otherwise be underestimated in a typical case (Graham, 2009; McKnight et al., 2007; Allison, 2003), similar to standard errors obtained by single imputation (cf. Sinharay et al., 2001). We therefore needed to decide over which MDT to apply (in place of EM).

Literature suggests to carry out hypothesis testing with MI or FIML (Graham, 2009, 2012d). Because (a) CFA parameters obtained in conjunction with FIML are generally unbiased if missings are MAR or MCAR (Enders, 2001b), because of (b) FIML's many advantages for most SEM applications (Allison, 2003), and because (c) a MI-based approach was not a viable option for our type of analysis (Graham, 2012e), we chose to apply the FIML procedure to our data next.

AMOS, our tool at hands for CFA and CBSEM, offers a preprocessing feature that–if activated–applies a FIML algorithm to a data set with missings while the actual analysis procedure (i. e., CFA, etc.) is performed.[22] However, this is only feasible if ML-based SEM estimation is selected. The model to be tested can be specified as usual using the graphical interface, but in addition to the actual model, the saturated and independence models must be fitted as well for FIML to be applied (Arbuckle, 2013); hence, the procedure requires to model or estimate the means and intercepts of the latent variables (Allison, 2003), even though one may not necessarily be interested in them as such (Weiber & Mühlhaus, 2010). For a single group analysis like ours, this is typically done by fixing the means and error terms of the factors to zero; labels for intercepts are not needed in this case (Byrne, 2010).

If configured as described above, all variables the model is comprised of will automatically be accounted for by the AMOS FIML algorithm, yet unless explicitly specified, for no further variables; Figure 4 shows the additional effort required to add a single auxiliary variable, namely *age*, to a simple first-order model with four correlated factors, here demonstrated for CIRME as an example (details are explained in Section 4.3 and Section 5.3). Our approach was therefore to apply the typical FIML strategy–thus an exclusive one (Graham, 2009)–when performing the various CFA and SEM required for our analysis.

4.6 VERIFYING FACTOR STRUCTURES AND RELIABILITY

At this stage the aim was to perform further reliability analyses and to verify the latent structures supported by previous analyses. A CFA tests the hypothesis that a scale is unidimensional and thus helps to establish or verify the dimensionality of a scale (Bagozzi & Yi, 2012). In order to correctly interpret obtained fit indexes and draw suitable inferences, it was also essential to evaluate the validity of these indexes in the context of our study. The next step therefore consisted of assessing the suitability of the various estimation methods and goodness-of-fit criteria available for CFA (thereby possibly anticipating the decision on the same subjects with regard to SEM) while particularly focusing on the characteristics of our data, as we discuss in the following.

4.6.1 *Choice of Estimation Method and Fit Measures*

It is important that data characteristics with regard to distributions match the requirements of the estimation method (cf. Kline, 2012), especially in order to interpreting the outcome of an analysis correctly, because the chosen estimation method has a strong impact on fit indexes (Fan, Thompson, & Wang, 1999).

22 If the data contains missings and this feature is not activated, AMOS will fail to run and exit with an corresponding error message.

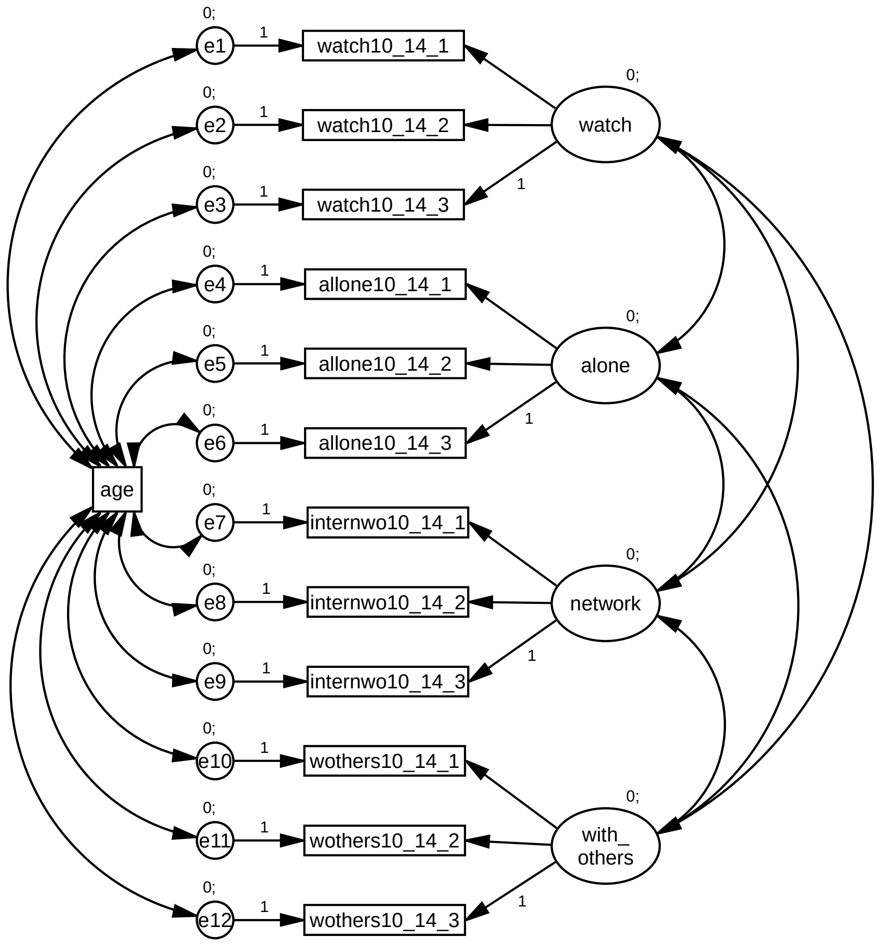


Figure 4: Input specification of a first-order childhood IR media experience (CIRME) model with four correlated factors before performing a ML-CFA on data with missing values with the aid of FIML and a single auxiliary variable, age. Latent factor means and error terms are fixed to zero for the application of FIML in AMOS; the model allows for correlations of age with the error terms of each measured endogenous variable of the model (cf. Allison, 2003, p. 550; note that there are no unobserved exogenous variables in this model).

ML estimation is considered relatively robust against departures from normality inasmuch as point estimates like loadings, variances, and covariances will not be affected (Bühner, 2011; Hancock & Liu, 2012; Finch, West, & MacKinnon, 1997) and as underestimated standard errors–a condition which may cause paths and correlation to appear significant although they are not–are not as much of a problem as often assumed (cf. Chou, Bentler, & Satorra, 1991). A yet important consequence of nonnormal data with regard to model fit, however, is a possible overestimation of the χ^2 statistics. The latter is an inferential test which is used as an overall fit measure for ML estimation, that is, to evaluate whether the model suits the data sufficiently. It inherently accounts for deviations of the analyzed data from a normal distribution, but notably also for sample size and model complexity (MacKenzie et al., 2011). If the χ^2 statistic for the data is artificially inflated, for example, due to the fact that the sample population is not normally distributed (Enders, 2001b), or because the sample is "adequately large" (MacCallum, Browne, & Sugawara, 1996, p. 132; cf. also Bagozzi & Yi, 2012; Bühner, 2011; Barrett, 2007), the discrepancy between the observed data and a tested model will be judged considerable on the grounds of a significant probability level for the χ^2 test (cf. MacKenzie et al., 2011). This will ultimately lead to the rejection of the model under investigation, although it might actually fit the data acceptably well (Bühner, 2011; Hancock & Liu, 2012; Bearden, Sharma, & Teel, 1982). (Hu, Bentler, & Kano, 1992) even warn against using the χ^2 goodness-of-fit test in general; in a study, they found that it is dramatically affected by the individual conditions present in an particular set of data. Some have suggested to use Hoelter's critical N for orientation (also provided by AMOS, cf. Arbuckle, 2013), as it indicates the largest sample size for which one would accept the hypothesis that a model is correct; due to its sensitivity to degrees of freedom (*df*), this recommendation yet appears highly questionable (Weiber & Mühlhaus, 2010).

One recommended approach to remedy this problem is to adjust estimates via bootstrapping χ^2 statistics (e. g., Byrne, 2010; West et al., 1995; Hancock & Liu, 2012), or to rescale the statistics by means of a correction factor (Satorra & Bentler, 1994; Hu et al., 1992; Chou et al., 1991; this option is currently not available in AMOS, cf. IBM AMOS Support, 2010). Other approaches involve renouncing the use of ML-based techniques and utilizing, for example, asymptotic distribution free (ADF) statistics instead (Ory & Mokhtarian, 2010; Chou et al., 1991; Browne, 1984). With regard to our study, on the one hand side, the ability to provide corrected χ^2 statistics in order to better adjudge the goodness of our model for CFA seemed to speak in favor of the bootstrap approach or the scaled χ^2 approach. On the other hand side, ADF had desirable qualities if a large sample size is available, which was the case with our data.

Unfortunately, the fact that we were not dealing with a complete data matrix complicated the matter again. First of all, common indexes of univariate or multivariate normality and of multivariate outliers cannot be calculated in the presence of missing data with software available today (a mentioned above, see Yuan et al., 2004, for experimental approaches in this direction), so important points of reference for deciding which estimation methods suits best or which multivariate outliers to remove were not available. Second, the recommended Bollen-Stine bootstrap cannot be performed (Bühner, 2011, p. 409; as for instance, AMOS does not provide any of the bootstrapping options otherwise available, cf. Arbuckle, 2013), and, as dealing with missings requires to estimate means and intercepts, only ML estimation can be requested (for example according to the corresponding error message in AMOS).[23] And third, besides the fact that the scaled χ^2 approach is only available in certain commercial programs, it cannot be applied to data with missings (Mplus Support, 2008).

We ultimately decided to perform the individual CFAs using ML estimation in conjunction with FIML (as demonstrated in Figure 4, but without auxiliary variables), thereby keeping in mind the questionability of significance levels obtained from the χ^2 tests both due to sample size and nonnormality (whereupon the impact of the latter factor was expected to be aggravated by the former).[24] In accord with the "common denominator" of several recommendations (Kline, 2011; MacKenzie et al., 2011; Bagozzi & Yi, 2012; Boomsma, Hoyle, & Panter, 2012; Beauducel & Wittmann, 2005), we report multiple fit measures, namely the root mean square error of approximation (RMSEA), the nonnormed fit index (NNFI) or Tucker-Lewis index (TLI), the two of which are interchangeable (Bagozzi, 2011), the comparative fit index (CFI), and the standardized root mean square residual (SRMR) as well as related figures and model properties for the CFAs performed. Note that the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the so-called normed χ^2, that is, χ^2/df, and the normed fit index (NFI) are not included due to criticism regarding their meaningfulness (cf. Bagozzi & Yi, 2012; Kline, 2011; Bühner, 2011). The χ^2 statistic is reported for orientation; however, with large sample sizes such as ours, large values are common (cf. Tickle, Hull, Sargent, Dalton, & Heatherton, 2006).

In AMOS, the SRMR cannot be calculated directly in the presence of missing values, because it is calculated from sample and residual moments; the situation should be the same with other tools. To obtain this index, we needed to follow a complicated work-around procedure which uses the FIML covariance matrix for the saturated model as input data computing the SRMR (IBM AMOS Support, 2012b, as demonstrated in Figure 5), that is, for each construct and set of indicators separately.

4.6.2 *Reliability, Fit, Sample Size, and Statistical Power*

As explained above, we intended to obtain further reliability indexes with the aid of CFA, for the purpose of validating the dimensional structure of our measures and also for comparing plausible competing (and possibly equivalent) solutions (cf. Bagozzi & Yi, 2012; Bühner, 2011; MacKenzie et al., 2011; Weiber & Mühlhaus, 2010; Floyd & Widaman,

23 Related thereto, it would seem a somewhat doubtful practice to use a ML-based MDT like FIML first and then subsequently apply an estimation method that assumes no distribution at all (like ADF).

24 Note that these individual CFAs did not include the player level, the only variable not within the reasonable limits of normality in this study.

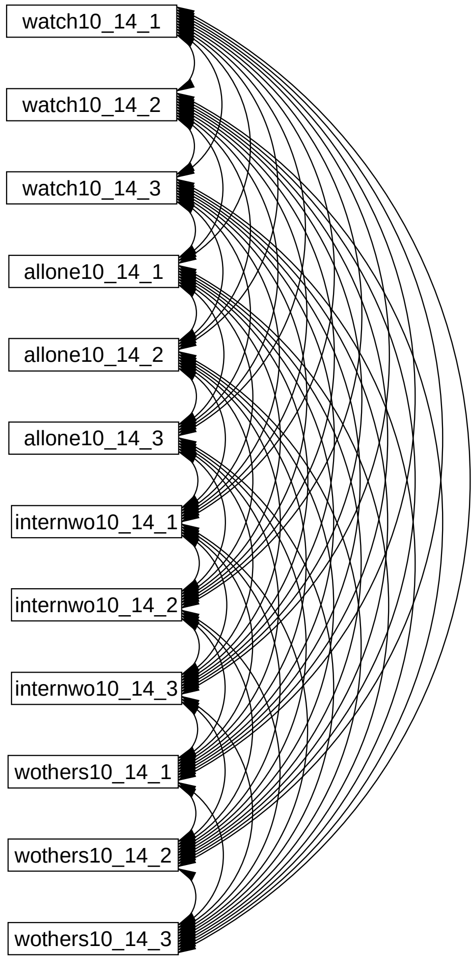


Figure 5: Input specification of the saturated childhood IR media experience (CIRME) model for SRMR computation with missing values in AMOS.

1995; T. A. Brown & Moore, 2012). The results of the exploratory analyses can be summarized as follows:

- The correct interpretation of IR emotional capabilities turned out to be the greatest challenge, because of, on the one hand side, several problems on the item level, and, on the other hand side, because of a certain lack of clarity with regard to factor structures. While we had assumed to find three factors, we initially found six. One of them was caused to vanished because reported similarities between certain items and their respective potential to form a distinctive factor were suspicious; the remaining indicators seemed to be best described by five factors, yet some of them were found to be problematic. After excluding said indicators for the simultaneous EFA, either assuming five or four factors, respectively, seemed appropriate, depending on whether 10 or 11 or rather 12 factors were to be extracted.

- Self-estimated cognitive ability had been hypothesized as a unidimensional construct, but the individual EFAs indicated the presence of a possible second factor. Results of the simultaneous EFA depended on how many factors were extracted: In a structure with 10 overall factors, the construct seemed to be best captured as unidimensional, yet with 11 and 12 factors, it was found to be two-dimensional.

- Different than postulated, cognitive absorption was found to be unidimensional most of the time (see initial PA), and after eliminating its unreliable indicator, all analyses pointed towards a unidimensional factor structure.

- The assumed latent four-factor structure of CIRME had not been contradicted by any of the analyses so far.

- Results for self-motivational traits consistently imposed a two-factor structure, and parental control, IR mediation, and IR enjoyment had been classified as unidimensional throughout all three of the previous analyses.

- The third factor of competitiveness suffered from the fact that one of its indicators exhibited significant cross-loadings at different occasions as well as reliability issues; one possible solution was therefore to drop one dimension, as only two "good" items remained that loaded on the latter.

We first cross-validated the results of the exploratory analyses for each construct via CFA (cf. Byrne, 2010, p. 164), thereby comparing the solutions implied by individual EFAs and the simultaneous EFAs. Furthermore, to finally decide whether to retain a second-order rather than a first-order structure for a particular construct, we estimated correlated factor models and their higher-order counterparts if their respective constructs were conceptually eligible for this type of comparison–that is, if neither a simple nor a higher-order factor structure was imposed by its conceptualization (on testing hierarchically nested models, see e. g., Kline, 2011, p. 214; F. F. Chen et al., 2005; on issues related to testing model nesting and equivalence, see Bentler & Satorra, 2010). Both corresponding models were then compared with respect to the scored index values, and the potential first-order factors were analyzed with regard to their correlations (for how to decide whether criteria for a second-order factor are sufficiently met, see Bagozzi & Yi, 2012, pp. 24–25).

As far as the analyses results are concerned, none of the variance estimates were negative and all procedures converged (cf. MacKenzie et al., 2011). The χ^2 statistics (i. e., uncorrected) were all significant

($p < .000$). Since most of our measurement models were standard CFA models, identification had not been expected to be of concern (cf. Kline, 2011, p. 137 et seqq.); the analyses confirmed this assumption. We then calculated the recommended fit index values for all constructs (χ^2 statistics, RMSEA $\leq .06$, NNFI $\geq .95$, CFI $\geq .95$, and SRMR $\leq .08$; these values are all very conservative, cf. Bagozzi & Yi, 2012 and MacKenzie et al., 2011). We found that all first-order model had obtained comparatively good fit (see Appendix D, Section D.2.2), and that none of the second-order models had achieved the highest fit index for any of the constructs. As fit and model selection indexes often penalize model complexity (West, Taylor, & Wu, 2012), this was yet not unexpected.[25]

As discussed earlier, hierarchical models of intelligence are well established (Bühner, 2011). However, as we only found support for two factors of self-estimated cognitive ability, a higher-order solution was not eligible (Weiber & Mühlhaus, 2010, p. 220; Kline, 2011, p. 249). The same was true of the emotion constructs, which were often found to be two-dimensional, yet not three-dimensional. On the basis of theoretical considerations (see Section 2.5) and the results obtained for the several first-order and second-order models, we concluded that–for the purpose of building strong theory–our overall model would not benefit from higher-order constructs.

For cognitive absorption, the correlated two-factor model and the one-factor model fitted equally well when comparing all obtained fit indexes; we thus needed further analysis to decide which one of them to prefer. In the case of competitiveness, various models yielded comparably good overall fit, too. On grounds of the χ^2 statistic, which clearly tended toward the more parsimonious model, we decided to drop the third dimension of personal standards in comparison with others.

On the basis of the respective indicator sets that were considered final according to their psychometric properties, the best fitting models of the primary constructs were then further assessed (cf. MacKenzie et al., 2011; Bagozzi & Yi, 2012; Kline, 2011; Weiber & Mühlhaus, 2010; Fornell & Larcker, 1981; Raykov, 2004; for issues with regard to using point estimates, related confidence intervals, and fixed cutoff values, and to solely relying on a single fit index, see F. Chen, Curran, Bollen, Kirby, & Paxton, 2008). Those assessments (first-order only, as second-

25 Note: After performing the analyses listed above, we noticed that 171 entries in our data to exhibit a standard deviation of 0 on all primary construct items, due to the fact that all of these items had received the same score (typically a “0”, which corresponded to a neutral score across all survey items). This indicated that the corresponding participants had not been motivated to give meaningful answers or to participate actively in the survey, and we preferred to exclude these entries from further analysis. We therefore examined whether their amount–in comparison with the overall sample–required further consideration in terms of distortion of prior analyses. We tested this by comparing CFA results that included said entries with CFA results that excluded them. Our findings were reassuring: If we detected a difference at all, it was negligible. The following sections report the results of those analyses which excluded these entries, thus including 4,048 respondents overall.

order did not apply) comprised the investigation of indexes known as the aforementioned SMCs and computed as[26]

$$\rho_i = \frac{\lambda_{ij}^2 \phi_j}{\lambda_{ij}^2 \phi_j + \theta_i} \ ,$$

where

λ_{ij}	loading of indicator i on its hypothesized factor ξ_j
ϕ_j	variance of ξ_j
θ_i	variance of the measurement error associated with i

and used as an index of indicator reliability. Furthermore, we calculated an additional index of factor reliability, often called composite reliability and computed as

$$\rho(\xi_j) = \frac{\left(\sum_{i=1}^k \lambda_{ij}\right)^2 \phi_j}{\left(\sum_{i=1}^k \lambda_{ij}\right)^2 + \sum_{i=1}^k \theta_i} \ ,$$

that is, for all $i = 1, 2, \ldots, k$ connected to the same ξ_j. Finally, we obtained the average variance extracted (AVE) of every factor, obtained by

$$\mathrm{AVE}(\xi_j) = \frac{\left(\sum_{i=1}^k \lambda_{ij}^2\right) \phi_j}{\left(\sum_{i=1}^k \lambda_{ij}^2\right) + \sum_{i=1}^k \theta_i} \ .$$

the latter being an index of convergent factor or construct validity which measures the amount of variance captured by a construct in relation to the variance due to random measurement error.[27] Tables with detailed results of these investigations are reported in Appendix D, Section D.2.2; however, only the best are shown.

As expected, the z-test of all estimates' critical ratios indicated statistical significance ($p < .001$, two-tailed) of the hypothesized relationships between indicators and factors (MacKenzie et al., 2011), factors and factors (with the exception of, rather surprisingly, act on others' emotions and regulate own emotions), as well as of variances and intercepts. All composite reliability indexes, as recommended, had values of $\geq .70$, and all AVE values were $\geq .50$ (cf. MacKenzie et al., 2011; Bagozzi & Yi, 2012) with the one exception of focused immersion (.492, i. e., of the two-factor model of cognitive absorption).

We also screened our results in order to check for unreliable indicator. We thereby gave removing the latter careful thought if this did not result in an improvement of the measurement model or if, by that

26 Note that multiple variations of the following notations exists; for comparison, see the aforementioned references. Also, shorter versions of the formulas exist for standardized values, see e. g., Bagozzi and Yi (1988), Weiber and Mühlhaus (2010).

27 Alternatively, it can be computed by "averaging the squared completely standardized factor loadings (λ^2) for the indicators, or by averaging the squared multiple correlations for the indicators" (MacKenzie et al., 2011, p. 313).

action, the number of indicators fell below the amount of three considered critical for estimation (cf. Bagozzi & Yi, 2012, p. 17, with regard to removing indicators solely based on their individual reliability values). Two of the primary construct items were significantly below the recommended .04 regarding their SMC value (cf. Weiber & Mühlhaus, 2010), that is, *cogabs_*7 (.312) of cognitive absorption and *watch*10_14_3 (.301) of the CIRME construct. For comparison, we recalculated composite reliability and AVEs after excluding the problematic indicators. For the watching IR action passively factor, this improved both values; yet as the factor only had three indicators, dropping one only left two indicators for its estimation. In the case of cognitive absorption, the picture was mixed. For the one-factor model, excluding *cogabs_*7 improved both values; for the two-factor model, though the AVE value of focused immersion improved by that, composite reliability slightly changed for the worse.

Results for the discriminant factor validity criterion developed by Fornell and Larcker (1981, cf. also MacKenzie et al., 2011, Weiber & Mühlhaus, 2010), which requests the squared correlation Φ_{ij}^2 of two factors ξ_i and ξ_j to be smaller than the AVE of each of those factors and thus is calculated as

$$\text{AVE}(\xi_i) \geq \Phi_{ij}^2 \qquad \text{for all } i \neq j,$$

were all satisfactory except for the two-factor model of cognitive absorption, for which the factors correlated so highly ($\Phi_{12} = .961, \Phi_{12}^2 = .924$; see Byrne, 2010, p. 168; see, Malone & Lubansky, 2012, p. 267 et seqq. on the topic of multicollinearity) that we had to reject the one-factor model (cf. Byrne, 2010, pp. 170–171, Kline, 2011, p. 72, and Bagozzi & Yi, 2012, p. 16). The one-factor model had generated desirable AVE (.534) and very good composite reliability (.888).[28] Additionally eliminating *cogabs_*7, which had exhibited little reliability in both models, improved the AVE (.572), so we excluded this item. Corresponding figures and the abovementioned fit indexes for each construct are presented in Appendix D, Section D.2, where the actual models are contrasted with their respective null (or independent) model, that is, a model with the same subset of variables assuming no relationships between any of them. The final item subsets were thereby formed with respect to prior results (see above). For comparison and to confirm these results, models were also tested with subsets that included items which were not found to be sufficiently reliable, yet to avoid confusion they are not displayed. The overall measurement model including all 12 factors and the 47 final indicators provided good overall fit (RMSEA: .041, NNFI: .906, CFI: .920, and SRMR: .0449), with the confidence intervals around RMSEA allowing rejection of an hypothesis of mediocre fit ($.040 \leq \epsilon \leq .042$).

28 Kline (2012) actually even states that “if a single-factor model cannot be rejected, there is little point in evaluating more complex ones” (p. 234).

Finally, we ran a common method bias analysis comparing the standard regression weights of a common latent factor solution with results of a solution without such a factor (cf. P. M. Podsakoff, MacKenzie, Jeong-Yeon Lee, & Podsakoff, 2003; MacKenzie & Podsakoff, 2012; Doty & Glick, 1998; Bagozzi, 2011). The highest difference between the standard regression weights of relationships between the hypothesized factors and their indicators across both solutions was .128 (after excluding *watch*10_14_3 and *cogabs*_7), so that no problems were to be expected in this regard.

To then evaluate the actual statistical power of the individual CFAs in combination with our sample size, we used the approach proposed by MacCallum et al. (1996), which determines the statistical power of a CBSEM analysis on the basis of the RMSEA and determines effect size by the difference between a null RMSEA value ϵ_0 and an alternative RMSEA value ϵ_a. To obtain a close fit, we set ϵ_0 to .05 and ϵ_a to .08 (cf. also Kline, 2011).[29] We applied an α-level of .01 per hypothesis as well as a β-level of .01 (the latter resulting in a power specification of .99), as our actual sample size of 4,048 was sufficiently large to satisfy these demanding requirements for most individual models (cf. Cohen, 1992). We chose to indicate the minimal sample size and to keep α and β at a fixed level in order to facilitate comparison; as the algorithm invariably output a power of 1 across various calculations due to limited precision, the performance measures would have been indistinguishable otherwise. Given the aforementioned parameters, the minimal sample sizes calculated for each individual model of our primary constructs ranged from 169 to 3,150. Required sample sizes for self-motivational traits and competitiveness ranged from 475 to 1,950, yet for parental control ($df = 5$), IR mediation ($df = 5$), and IR enjoyment ($df = 2$), the available sample size could only account for a statistical power of .97 and .64, respectively (i. e., ceteris paribus). Figures are displayed in Appendix D, Section D.2. The overall measurement model was shown to require a minimal sample size of 70 under the same conditions.

With regard to IR emotional capabilities, we found support for five remaining factors in our data (in parentheses the corresponding temporary labels of the factors from the CFA results), namely (a) perceiving others' IR emotions (= F1), (b) perceiving / understanding own IR emotions (= F4), (c) understanding others' IR emotions (= F5), (d) regulating own IR emotions (= F3), and (e) acting on others' IR emotions (= F2). In alignment with our hypotheses, the five factors of IR emotional capabilities belonged to the distinct constructs of perceive IR emotions, understand IR emotions, and manage IR emotions. However, the factor of perceiving own IR emotions had essentially vanished at this point, while the item left of this factor rather seemed to represent an item of the understanding own IR emotions factor. Also, as discussed

29 See MacCallum et al., 1996, pp. 132–135 for reasons why a close fit is to be preferred over an exact fit here.

above, though a higher-order approach would have been justifiable for perceiving and managing IR emotions, two-dimensional constructs can yet not be modeled as a higher-order construct. Factor correlations obtained through the various CFAs we performed also supported the hypothesized relationships between many of the IR emotional capabilities factors. It yet also became apparent that our assumption that regulating own IR emotions and acting on others' IR emotions belonged to the same subconstruct, conceptualized as managing IR emotions, could not be sustained; the covariance (.12) between both factors turned out to be not significant ($p_{cov} = .006$), while all other covariances between IR emotional capability factors were significant (at the .001 level).

In order to provide a more meaningful model, we needed to make slight adjustments to the original research model regarding the causality chain of IR emotional capabilities, particularly concerning the compounds of perceiving, understanding, and managing IR emotions, as can be seen in the next section. The primary construct factors for which–after completing all exploratory analysis steps and individual CFAs–the best support was found and which qualified to be kept throughout the whole analysis process are displayed in Table 15, as well as their final interpretation.[30]

Specifically, we needed to reconsider the role of IR emotion regulation and acting on others' IR emotions as well as their interplay with other IR emotional capabilities, yet also the causality chain of emotion-related events hypothesized by Joseph and Newman (2010). As their perception construct does not distinguish between perceiving own and perceiving others' emotions and the understanding construct does not distinguish between understanding own and others' emotions (cf. e. g., Joseph & Newman, 2010), we had to assume that both understanding factors would be affected by perceiving others' IR emotions. At the same time, understanding own IR emotions should precede understanding others' IR emotions as well as regulate own IR emotions. By the same token, understanding others' IR emotions should influence the regulation of own IR emotions as well as acting on others' IR emotions. The assumed relationships between IR emotional capabilities and their hypothesized predictors were adapted in a similar fashion. For example, analysis of the SE cognitive ability construct pointed to a two-factor structure, while the covariance between watching IR action passively and experiencing IR with others via network turned out to be not significant when validating the full measurement model. The final model we tested is depicted in the next section, where we also explain the adaptation of relationships between constructs in more detail.

30 In this context, Kline (2011) cautions against committing what he calls the "naming fallacy" (p. 230 and p. 365), namely to believe that a hypothetical construct or factor is correctly understood or labeled just because it has been given a name.

Table 15: Dimensionality of primary constructs supported by CFA

Construct(s)	Factors	Interpretation of Factors
IR Emotional Capabilities	5	(a) Perceive others' IR emotions (= F1), (b) understand own IR emotions (= F4), (c) understand others' IR emotions (= F5), (d) regulate own IR emotions (= F3), and (e) act on others' IR emotions (= F2)
SE Cognitive Ability	2	(a) Quotidian achievement feedback and (b) institutional achievement feedback
Cognitive Absorption	1	Unidimensional, as discriminant validity for two factors was not sufficiently supported
CIRME	4	(a) Watch IR action passively, (b) experience IR on one's own and alone, (c) experience IR with others via network, and (d) experience IR with others in physical proximity

4.7 HYPOTHESES TESTING

As a consequence of the restructuring of our measurement model, a revision of our hypotheses and our research model became necessary (Byrne, 2010). The latter was replaced by the model depicted in Figure 6[31], which reflects the following adjustments (see also Table 15):

- Regulating own IR emotions and acting on others' IR emotions affect IR performance: H1 replaced by H1a and H1b;
- understanding one's own IR emotions affects the regulation of one's own IR emotions and the understand of others' IR emotions, while understanding others' IR emotions affects regulating one's own IR emotions and acting on others' IR emotions: H2 replaced by H2a, H2b, H2c, and H2d;
- perceiving others' IR emotions affects understanding own and others' IR emotions: H3 replaced by H3a and H3b;

31 Note that for reasons of complexity hypothesis H1m is not part of the revised figure; the results of the multigroup analysis are explained in detail at the end of this chapter.

- both cognitive ability dimensions affect IR performance as well as both understanding IR emotions facets: additional hypotheses H4c, H4d, H4e, and H4f;
- cognitive absorption affects IR performance as well as perceiving others' IR emotions and both understanding IR emotion facets: adjusted hypotheses H5a, H5b, and H5c, additional hypothesis H5d;
- watching IR action passively affects perceiving and understanding others' IR emotions: adjusted hypotheses H6a and H6b;
- experiencing IR alone affects understanding own IR emotions: adjusted hypotheses H6d; note that H6c has been excluded due to changes in the perceiving construct;
- experiencing IR with others via network affects regulating own and acting on others' IR emotions: adjusted hypothesis H6f, additional hypothesis H6j; and finally,
- experiencing IR with others in close physical proximity affects regulating own and acting on others' IR emotions: adjusted hypothesis H6h, additional hypothesis H6k.

As aforementioned, performance was measured by a single indicator, the manifest variable *esllevel*. It had thus not been included in the measurement investigation activities (examination of scales, EFAs, etc.) presented above. We considered this single-item construct to measure without measurement error, hence we chose to fix its variance to zero in our full SEM (cf. e. g., Weiber & Mühlhaus, 2010, p. 153, Kline, 2011, p. 118, and Hoyle, 2012c, p. 132). We also considered its construct indicator to measure on a ratio or interval scale basis, respectively (Weiber & Mühlhaus, 2010). The actual variable data, however, had been found to exceed the range of normality to a nonignorable extend (cf. Figure B.1, Appendix B, Section B.2). We suspected that under these circumstances, an ML estimation could possibly not be as well-suited for our analysis as we would wish: When item distributions differ, categorization effects may occur (Bernstein & Teng, 1989), and additionally, differing metrics may lead to serious estimation problems (Bühner, 2011, p. 418). In an attempt to further sustain results obtained by ML-based CBSEM, we searched for a complementary approach in order to be able to control for distribution effects of the performance variable *esllevel* as well as the slight deviance from a multivariate normal distribution observed in the data as a whole. However, an estimation on the basis of the PLS technique, often recommended as a remedy to distribution issues, is not less error-proof than ML-based CBSEM when variables depart from a normal distribution (Goodhue, Lewis, & Thompson, 2012, see also Section 5.3). It has also been seriously doubted whether the PLS algorithm is able to cope with missing

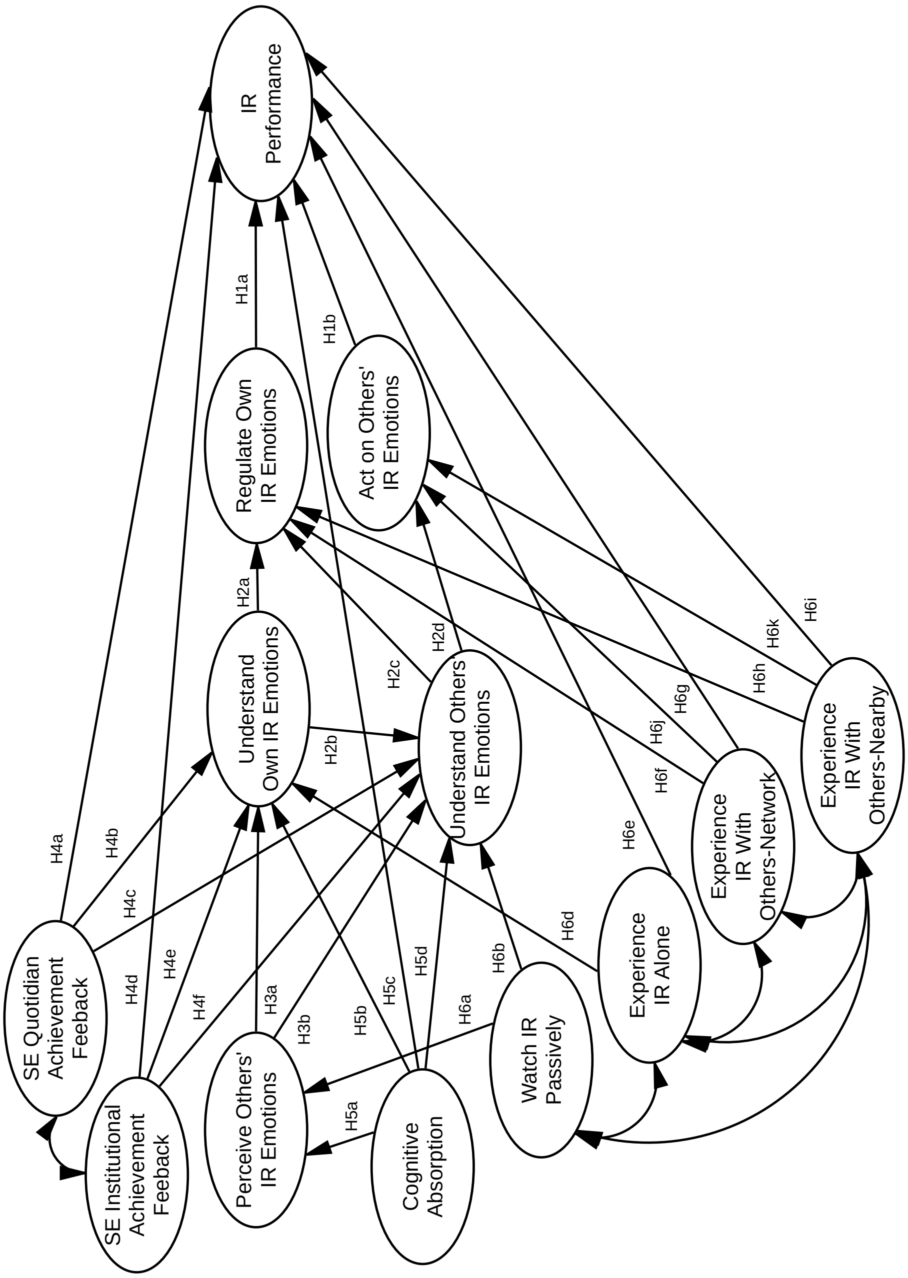


Figure 6: Revised hypothesized structural model of IR performance; hypothesis H1m not depicted.

values such that proper results are to be expected (Parwoll & Wagner, 2012). Other estimation methods like ADF produce results which should not be compared to ML estimates (Bühner, 2011)–even if otherwise, they were not applicable to our study due to missing values (see above).

Our final approach to this issue was threefold. First, we ran a regular ML-based CBSEM analysis on the basis of the original data. Second, to account for deviations of our performance measure, we decided to perform a data transformation for the purpose of obtaining an approximately normally distributed data of the original performance variable (*esllevel*: N = 4,048, M = 31.55 (.541), SD = 34.44, skewness: 3.215, kurtosis: 17.719). In this vein, we chose to apply a so-called Box-Cox transformation, which is explicitly suitable for a dependent variable such as our performance measure (Box & Cox, 1964; IBM AMOS Support, 2012a). The transformation can be applied to data through a web tool, for example.[32] Unfortunately, as one of the maximum correlations turned out to be negative–a condition which leads to unexpected results–a Box-Cox transformation was not feasible with our data. Our final solution consisted of forming seven performance categories which mirrored the response format of the other items and assigning participants to the resulting seven groups according to the following pattern: The first performance category contained all participants with a player level of 0, and the category with the highest player levels received approximately the same amount of participants as the first. The remaining five categories were grouped in such a way that the overall distribution resembled a normal distribution. This transformation (*esltrans*: N = 4,048, M = 3.98 (.024), SD = 1.506, skewness: .028, kurtosis: -.407) inherently solved any issues related to player level outliers and differing scale levels across different measurements and allowed for an ML estimation. We then re-estimated the model, this time using the transformed performance measure.

Third, in addition to our two CBSEM estimations, we also estimated our model using Bayesian structural equation modeling (BSEM), an approach which performs a SEM estimation with the aid of Markov chain Monte Carlo (MCMC) techniques (cf. D. Kaplan & Depaoli, 2012; Bakker, 2009; Dunson, Palomo, & Bollen, 2005; MacCallum, Edwards, & Cai, 2012; J. Gill, 2004). This method of analysis has been used for the triangulation of ML results before (e. g., Arbuckle, 2013), for example, in election studies (Bakker, 2009; cf. also Palomo, Dunson, & Bollen, 2007; Dunson et al., 2005; J. Gill, 2004). The approach has several advantages, particularly with regard to distribution assumptions (MacCallum et al., 2012; D. Kaplan & Depaoli, 2012), and is available in AMOS (instructions how to compare ML and BSEM are given

32 Wessa P., (2013), Box-Cox Normality Plot (v1.1.5) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL: http://www.wessa.net/rwasp_boxcoxnorm.wasp/.

by Arbuckle, 2013). In AMOS, BSEM requires to estimate means and intercepts, thus missing values are handled as part of the estimation process (cf. Arbuckle, 2013). The MCMC algorithm used for BSEM by AMOS draws random values of parameters from high-dimensional joint posterior distributions via Monte Carlo simulation of the posterior distribution of parameters. As our sample had a considerable size, we believed the influence of setting a prior distribution to be very small, hence we left the default setting of AMOS for Bayesian analysis unchanged. Furthermore, we applied a conservative convergence criterion of 1.002 (i. e., close to perfect convergence) and a very conservative burn-in period of 6,500 sample draws (cf. Arbuckle, 2013).[33] AMOS ran 32 batches with 11,905 and thereby generated approximately 381,000 observations in total. The results of a Bayesian analysis in AMOS returns various values which describe the marginal posterior distribution of a single model parameter. The posterior mean, that is, the center or average of the posterior distribution based on the data and the prior distribution, is typically used as the Bayesian point estimate of the parameter; is not the exact posterior mean but is an estimate obtained by averaging across the random samples produced by the MCMC procedure.

We applied BSEM to the untransformed data in order to compare its results with those of the two ML-based CBSEM estimations (i. e., one including the untransformed performance data, the other including the transformed performance data). Results of all three estimations are presented in Table 16. It is important to note that when using BSEM as opposed to ML CBSEM, the reported estimated standard error does not represent an estimate of how far the posterior mean may lie from the unknown true value of the parameter, but rather suggests how far the Monte-Carlo estimated posterior mean may lie from the true posterior mean. With the MCMC procedure continuing and generating more samples, the posterior mean becomes more precise, and the value of the standard error will gradually drop (Arbuckle, 2013). However, the likely distance between the posterior mean and the unknown true parameter is reported via the standard deviation of the estimation, which is analogous to the standard error in ML estimation (Arbuckle, 2013). Consequently, we display the standard errors for the two ML estimations and the estimates of deviation for the Bayesian analysis. According to both ML estimations, coefficients were significant if greater than .08 ($p < .001$). A table displaying the intercorrelations as well as the means and standard deviations of all variables included in the analysis (untransformed, ML estimation) can be found in Appendix D, Section D.3.[34]

33 Setting a high number of burn-in values also helps to avoid issues related with system memory which easily occur when AMOS calculates the estimates.

34 By reporting these particular figures, we follow the recommendations of the *Publication manual of the American Psychological Association* (2011, p. 133). Note that some authors rather present the covariance matrix of indicators instead (cf. Bagozzi & Yi, 2012).

Table 16: Standardized coefficients and standard errors / deviations for all relationships between constructs in the revised research model

Relationship			not transf.	transf.	Bayesian
F1	—>	F4	.33 (.021)	.33 (.021)	.33 [.021]
F1	—>	F5	.53 (.018)	.53 (.018)	.53 [.019]
F2	—>	PERF	.08 (.462)	.09 (.020)	.08 [.021]
F3	—>	PERF	-.01 (.613)	-.01 (.028)	-.01 [.022]
F4	—>	F3	.30 (.027)	.30 (.027)	.31 [.026]
F4	—>	F5	.19 (.020)	.19 (.020)	.19 [.022]
F5	—>	F2	.38 (.028)	.38 (.028)	.38 [.019]
F5	—>	F3	.15 (.028)	.15 (.028)	.15 [.026]
IAF	—>	F4	.32 (.036)	.32 (.036)	.31 [.032]
IAF	—>	F5	.17 (.028)	.17 (.028)	.17 [.029]
IAF	—>	PERF	-.05 (.939)	.00 (.041)	-.05 [.030]
QAF	—>	F4	-.02 (.029)	-.02 (.029)	-.02 [.029]
QAF	—>	F5	.01 (.022)	.01 (.022)	.01 [.025]
QAF	—>	PERF	.04 (.761)	-.01 (0.34)	.04 [.028]
CA	—>	F1	.32 (.023)	.32 (.023)	.32 [.019]
CA	—>	F4	.16 (.024)	.16 (.024)	.16 [.023]
CA	—>	F5	.08 (.018)	.08 (.018)	.08 [.019]
CA	—>	PERF	.00 (.678)	.01 (.030)	.00 [.024]
WP	—>	F1	.14 (.025)	.14 (.025)	.14 [.025]
WP	—>	F5	.05 (.017)	.05 (.017)	.05 [.019]
AL	—>	F4	-.01 (.019)	-.01 (.019)	-.01 [.023]
AL	—>	PERF	-.02 (.553)	-.03 (.023)	-.02 [.024]
NET	—>	F2	.12 (.018)	.12 (.018)	.12 [.022]
NET	—>	F3	.08 (.016)	.08 (.016)	.08 [.025]
NET	—>	PERF	.03 (.403)	.03 (.016)	.03 [.022]
WO	—>	F2	-.01 (.027)	-.01 (.027)	-.01 [.024]
WO	—>	F3	.02 (.024)	.02 (.024)	.02 [.027]
WO	—>	PERF	-.02 (.642)	-.03 (.026)	-.02 [.025]

Note. transf. = transformed performance data. ML standard errors in parentheses, Bayesian standardized deviations of estimation in brackets. Coefficients greater than .08 are significant at the $p < .001$ level according to ML estimation. CA = cognitive absorption, F1 = perceive others' IR emotions, WP = watch IR action passively, F4 = understand own IR emotions, AL = experience IR alone, IAF = institutional achievement feedback, QAF = quotidian achievement feedback, F5 = understand others' IR emotions, F3 = regulate own IR emotions, F2 = act on others' IR emotions, WO = experience IR with others in physical proximity, NET = experience IR with others via network, PERF = IR performance.

As can be taken from Table 16, for estimations which did not involve performance, the standardized coefficients were identical for all three estimation types (with the exception of F4 –> F3, for which the coefficient is .30 for both ML estimates and .31 for BSEM). This is not surprising, because in large datasets, the posterior mean of BSEM estimates tends to be close to the ML estimates (Arbuckle, 2013). Also, the standard errors (or standard deviations of the estimation, respectively) were relatively low and quite homogeneous; moreover, they all lay in a range between .017 and .036 (cf. Weiber & Mühlhaus, 2010, p. 182). In contrast, notable differences can be observed when looking at estimations that did involve performance. While the standardized coefficients of the nontransformed ML analysis and the Bayesian analysis were identical, they sometimes differed distinctly when compared with the results of the transformed ML analysis. On the other hand, standard errors of the nontransformed ML analysis were considerably higher (.403 to .939) than for the other two analyses (i. e., deviation of estimation in the case of BSEM), where these values were–again–relatively low as well as homogeneous and ranged between .016 and .041 (though all slightly more homogeneous in the case of BSEM).

Both ML estimations obtained very similar and satisfying goodness-of-fit indexes. The RMSEA of .042–with confidence intervals around RMSEA of $.041 \leq \epsilon \leq .043$–was even below the strictest recommended value of .05. The NNFI of .898 (untransformed performance data) and .899 (transformed performance data), respectively, as well as the CFI of .911 were very close to the cut-off value of .90 suggested previously by some (cf. Bagozzi & Dholakia, 2006; Nah et al., 2011; Weiber & Mühlhaus, 2010). The SRMR of .074 was also below the recommended value of .08. Hence, based on these indexes, our model had very good fit. The minimal sample size to obtain close fit was 68.

However, the results of the Bayesian analysis seemed to provide the most stable and trustworthy solution overall; Figure 7 depicts its results with the aid of our revised model. The BSEM obtained a posterior predictive p of .50, thus an excellent fit measure. As we do not compare different models, the obtained deviance information criterion (DIC) is not meaningful here (cf. also Spiegelhalter, Best, Carlin, & van der Linde, Angelika, 2002; Gelman, 2013). Examples of diagnostic plots provided by AMOS to check the convergence of the Bayesian analysis are displayed in Appendix D, Section D.3.2. The results of our hypothesis testing–summarized in Table 17–can be described as follows:

- H1a (regulate own IR emotions –> IR performance) is not supported, as the effect is very small and in the opposite direction than hypothesized. H1b (act on others' IR emotions –> IR performance) is supported, indicating that acting on others' IR emotions contributes to performance.

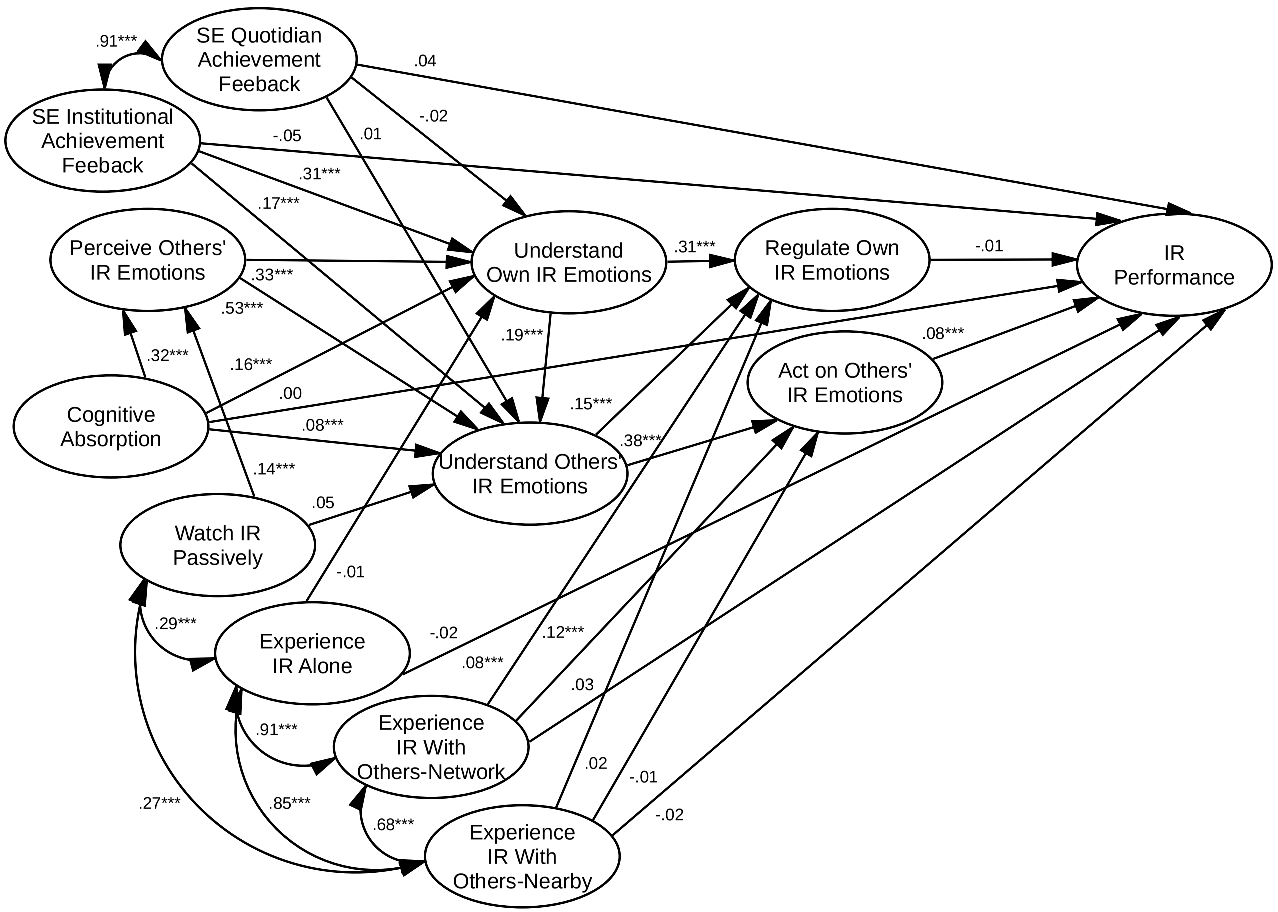


Figure 7: Results of testing the revised IR performance model with the observed data (N = 4,048, $^{***}p < .001$ according to ML estimation).

- H2a (understand own IR emotions –> regulate own IR emotions), H2b (understand own IR emotions –> understand others' IR emotions), H2c (understand others' IR emotions –> regulate own IR emotions) as well as H2d (understand others' IR emotions –> act on others' IR emotions) are supported, suggesting an effect of understanding emotions on an emotional response to these emotions in an IR context.
- H3a (perceive others' IR emotions –> understand own IR emotions) and H3b (perceive others' IR emotions –> understand others' IR emotions) are supported, indicating that perceiving emotions precedes understanding emotions in an IR context.
- H4a (quotidian achievement feedback –> IR performance), H4b (quotidian achievement feedback –> understand own IR emotions), H4c (quotidian achievement feedback –> understand others' IR emotions), H4d (institutional achievement feedback –> IR performance) are either not supported or did not prove to represent significant effects; however, H4e (institutional achievement feedback –> understand own IR emotions) and H4f (institutional achievement feedback –> understand others' IR emotions) are supported, indicating that SE cognitive ability of importance for institutional achievement also supports understanding IR emotions.
- H5a (cognitive absorption –> perceive others' IR emotions), H5b (cognitive absorption –> understand own IR emotions), and H5d (cognitive absorption –> understand others' IR emotions) are supported, suggesting that cognitive absorption affects various IR emotional capabilities; however, H5c (cognitive absorption –> IR performance) does not represent a significant effect.
- H6a (watch IR action passively –> perceive others' IR emotions) is supported indicating that this facet of CIRME contributes to the perception of others' emotions in an IR context, while H6b (watch IR action passively –> understand others' IR emotions) is not supported.
- H6d (experience IR alone –> understand own IR emotions) is not supported.[35]
- H6f (experience IR with others via network –> regulate own IR emotions) and H6j (experience IR with others via network –> IR performance) did not prove to represent significant effects.
- Finally, H6h (experience IR with others in close physical proximity –> regulate own IR emotions) is supported, whereas H6k (experience IR with others in close physical proximity –> acting on others' IR emotions) is not supported.

35 Note that H6c has been excluded due to changes in the perceiving IR emotions construct.

Table 17: Results of hypothesis testing

Hypothesis				Supported?
H1a	F3	—>	PERF	no
H1b	F2	—>	PERF	yes
H2a	F4	—>	F3	yes
H2b	F4	—>	F5	yes
H2c	F5	—>	F3	yes
H2d	F5	—>	F2	yes
H3a	F1	—>	F4	yes
H3b	F1	—>	F5	yes
H4a	QAF	—>	PERF	n.s.
H4b	QAF	—>	F4	no
H4c	QAF	—>	F5	n.s.
H4d	IAF	—>	PERF	no
H4e	IAF	—>	F4	yes
H4f	IAF	—>	F5	yes
H5a	CA	—>	F1	yes
H5b	CA	—>	F4	yes
H5c	CA	—>	PERF	n.s.
H5d	CA	—>	F5	yes
H6a	WP	—>	F1	yes
H6b	WP	—>	F5	n.s.
H6d	AL	—>	F4	no
H6e	AL	—>	PERF	no
H6f	NET	—>	F3	yes
H6g	NET	—>	PERF	n.s.
H6h	WO	—>	F3	n.s.
H6i	WO	—>	PERF	no
H6j	NET	—>	F2	yes
H6k	WO	—>	F2	no

Note. n.s. = not significant ($p < .001$). There is no H6c due to the revisions above. CA = cognitive absorption, F1 = perceive others' IR emotions, WP = watch IR action passively, F4 = understand own IR emotions, AL = experience IR alone, IAF = institutional achievement feedback, QAF = quotidian achievement feedback, F5 = understand others' IR emotions, F3 = regulate own IR emotions, F2 = act on others' IR emotions, WO = experience IR with others in physical proximity, NET = experience IR with others via network, PERF = IR performance.

According to both ML estimations, the model accounted for 13% of perceiving others' IR emotions, 31% of understanding own IR emotions, 54.7% of understanding others' IR emotions, 17.6% of regulating own IR emotions, 16.3% of acting on others' IR emotions, and 1% of IR performance. The BSEM output in AMOS does not provide comparable values. However, calculation by hand pointed to the same explanatory value of the exogenous variables for the endogenous variables obtained from using Bayesian analysis for the model.

We further tested our model for full mediation, in order to check whether direct paths not hypothesized but relevant to our research are nonsignificant (cf. Bagozzi & Yi, 2012; Bagozzi & Dholakia, 2006; Bergami & Bagozzi, 2000). Due to the fact that our BSEM analysis was very costly in time and system memory, we performed these tests using ML. We provide the obtained goodness-of-fit indexes for each model for orientation; the results are presented in Table 18. However, as elaborated above, the χ^2 is very sensitive to large sample sizes, hence a χ^2-difference-test was not applicable with respect to our data.

Table 18: Results of testing the hypothesized mediation

Model		Fit indexes	Test of hypothesis and conclusion
M1	Baseline (Figure 6)	$\chi^2(1032)$=8,413.86, RMSEA=.041, NNFI=.899, CFI=.911, SRMR=.074	Revised model in Figure 6 is consistent with the data.
M2	Added paths F1 —> F2, F1 —> F3	$\chi^2(1030)$=8,337.69, RMSEA=.042, NNFI=.899, CFI=.912, SRMR=.072	First added path significant (p<.001), second not (p=.969); model could benefit from an alteration according to χ^2, CFI, and SMRM (but not RMSEA or NNFI).
M3	Added path F1 —> PERF	$\chi^2(1031)$=12,326.83, RMSEA=.052, NNFI=.894, CFI=.903, SRMR=.074	Added path not significant (p=.203), model should not be altered (cf. also fit indexes).

Note. F1 = perceive others' IR emotions, F2 = act on others' IR emotions, F3 = regulate own IR emotions, PERF = IR performance.

Additionally, to investigate the potential situational moderation for emotional capability effects on performance (cf. hypothesis H1m), we conducted a moderator analysis similar to that of Joseph and Newman (2010) by performing a multigroup estimation (again using ML). To this end, we split our sample into two distinct subpopulations, one having a preference for virtual worlds with a shooter and fighter theme, the other having a preference for themes like strategy, social games, adventures, role playing, and so on (according to variable gm_genre_offer). It is important to note that the part of the group with no data on this variable was not included in this analysis; because the distinction between the two groups is crucial to the outcome of estimation, we decided to only include cases with no missings on this variable–rather than letting FIML estimating it. A multigroup analysis in AMOS requires to physically split the initial data file according to a selection variable so that the analysis can be performed for each group separately; however, if the data on that variable is missing, a case cannot be assigned to any of the groups (does the missing value indicate a preference for shooter or rather nonshooter virtual worlds?). As a result, the total sample size was reduced to 2,435 individuals (shooter sample: 1,385; nonshooter sample: 1,050), in contrast to the original sample with 4,048 cases. As accounting for a possible bias introduced through that selection is not feasible, the results have to be interpreted accordingly.[36]

The results of the multigroup analysis are shown in Table 19. While all effects were almost constant across the different groups (note the very strict p levels), a remarkable difference can be observed when comparing the estimate for the path connecting acting on others' IR emotions (F2) to performance (PERF). The size of this effect of .08 was highly significant for the complete sample; however, for the shooter sample it dropped to .05 and became insignificant, while for the nonshooter sample it was stronger (.14) than for the complete sample and highly significant ($p < .001$). This suggests that different preferences for genres of virtual worlds–which reflect the type of virtual worlds that the users under investigation engage in–moderates the effects of emotional capabilities with regard to performance. This also confirms findings of Joseph and Newman (2010) related to emotional labor in the context of IR.

In this chapter, we have outlined the preliminary data preparation, the investigation of data characteristics, the treatment of missing values, the examination of psychometric properties of measures, and the results of hypothesis testing and secondary analysis. In the upcoming chapter, we further interpret our results with respect to our hypotheses. We then address possible limitations of our study. Subsequently, we discuss the importance of missing values and other topics related to data issues and methods of analysis.

36 Joseph and Newman (2010) encountered similar issues (cf. p. 70).

Table 19: Standardized coefficients and corresponding probability levels for hypothesized paths when testing the revised model on data of the entire sample, the shooter group, and the nonshooter group

Relationship			Sample		Shooters		Nonshooters	
F1	—>	F4	.33	***	.39	***	.23	***
F1	—>	F5	.53	***	.56	***	.49	***
F2	—>	PERF	.08	***	.05	.101	.14	***
F3	—>	PERF	-.01	.676	-.04	.166	.02	.579
F4	—>	F3	.30	***	.33	***	.26	***
F4	—>	F5	.19	***	.20	***	.17	***
F5	—>	F2	.38	***	.38	***	.37	***
F5	—>	F3	.15	***	.13	***	.15	***
IAF	—>	F4	.32	***	.30	***	.39	***
IAF	—>	F5	.17	***	.20	***	.17	.001
IAF	—>	PERF	-.05	.100	.06	.263	-.18	.002
QAF	—>	F4	-.02	.452	-.06	.212	-.02	.728
QAF	—>	F5	.01	.790	.02	.637	.01	.802
QAF	—>	PERF	.04	.197	-.06	.217	.11	.041
CA	—>	F1	.32	***	.31	***	.31	***
CA	—>	F4	.16	***	.13	***	.20	***
CA	—>	F5	.08	***	.06	.026	.09	.010
CA	—>	PERF	.00	.914	.00	.952	.01	.736
WP	—>	F1	.14	***	.14	***	.17	***
WP	—>	F5	.05	.009	.05	.101	.06	.140
AL	—>	F4	-.01	.621	-.02	.621	.00	.965
AL	—>	PERF	-.02	.474	-.05	.227	-.07	.147
NET	—>	F2	.12	***	.12	***	.12	.003
NET	—>	F3	.08	***	.08	.023	.08	.061
NET	—>	PERF	.03	.203	.09	.021	.03	.434
WO	—>	F2	-.01	.771	-.06	.121	.06	.190
WO	—>	F3	.02	.534	.05	.148	.04	.344
WO	—>	PERF	-.02	.345	-.04	.360	-.06	.261

Note. CA = cognitive absorption, F1 = perceive others' IR emotions, WP = watch IR action passively, F4 = understand own IR emotions, AL = experience IR alone, IAF = institutional achievement feedback, QAF = quotidian achievement feedback, F5 = understand others' IR emotions, F3 = regulate own IR emotions, F2 = act on others' IR emotions, WO = experience IR with others in physical proximity, NET = experience IR with others via network, PERF = IR performance. *** = significant at the $p < .001$ level according to ML estimation.

5 DISCUSSION

Life is the art of drawing sufficient conclusions from insufficient premises.

— *Samuel Butler*

The last chapter has explained the course of the analysis and presented its results. This chapter now interprets these results, qualifies them and draws inferences and conclusions with regard to our initial hypotheses. It also attempts to identify and evaluate possible limitations or issues related to our study.

5.1 INTERPRETATION OF RESULTS

The current study sought to clarify the predictive power of selected individual differences–notably emotional capabilities, cognitive ability, cognitive absorption, and media literacy–for individual virtual world user performance. To this end, we searched for evidence for a causal chain between postulated emotion subconstructs. Also, due to its importance according to prior studies, we investigated to which degree cognitive ability defines performance and examined how cognitive ability relates to the understanding emotion facet. Another important question we sought to answer was whether different levels of emotional labor may affect the connection between emotional capabilities and performance in different ways. In this vein, we built on highly influential works from psychology and tested a model which included three theoretically and empirically sound subfacets of EI as well as SE cognitive ability. We thereby borrowed knowledge from a meta-analytic study and transferred it into a self-contained study.

Furthermore, by incorporating insights from IS and media research, the current study contributed to a better understanding of the role that cognitive absorption as well as experience with IR media during childhood and early adolescence play with regard to IR performance. We not only analyzed whether cognitive absorption impacts performance directly, but also investigated whether it serves as an amplifier for the emotion-related processes preceding performance. We were further interested in the question whether CIRME supports the formation of performance-related skills as well as the development of emotional capabilities in various ways.

First, our findings seem to confirm that there is evidence for subfacets of emotional capabilities which influence each other in a sequential relationship that can be characterized as a causal chain. Moreover, our results suggest that in the context of virtual worlds and IR, emotional capabilities appear to have a similar impact on performance as they do in real-world contexts. The results further indicate that SE cognitive ability, cognitive absorption, and various kinds of experiences with IR media during childhood influence multiple facets of emotional capabilities, which in turn have a varyingly strong effect among different groups on IR performance through acting on others' IR emotions. Notably, in the present study, the effect was larger for users which prefer virtual worlds that have more emotional content and require more require more social and strategic skills, for example in virtual worlds representing social or role playing games or strategic tasks related to human behavior. Second, only the facet of SE cognitive ability accounting for institutional feedback seemed to be of predictive value in our model, that is, in explaining understanding IR emotions, and counterintuitive to our expectations, neither SE cognitive ability facet seemed to be directly linked to IR performance. Third, we found evidence for the impact of cognitive absorption on perceiving and understanding IR emotions; however, a significant direct path to performance was not confirmed by our data. And fourth, watching IR passively and experiencing IR with others via network showed significant impact on IR emotion perception as well as on managing IR emotions; nonetheless, effects on IR performance were nonsignificant.

When interpreting these results, it is noteworthy that, according to our control items, using virtual worlds with others via network (thus with no one in close physical proximity) is by far the most typical use condition our survey candidates reported (91.8%). One can infer from this that experiences with situations which require certain regulative processes with respect to one's own emotions are of little relevance because they hardly ever occur. The fact that individuals in the present mostly study used virtual worlds alone and only connected to others via network may also indicate why no significant covariance between watching IR action passively and experiencing IR with others via network had been identified before and further why the situations accounted for by the remaining CIRME factors had such inconsistent effects regarding the predictive value of CIRME. Moreover, this substantiates our conclusion that the different types of tasks related to different types of virtual worlds may require different types of emotion regulation or management strategies.

The mechanisms of emotion regulation have recently been scrutinized by Webb, Miles, and Sheeran (2012). Their meta-analysis revealed that the selection process for human emotion regulation strategies can be rather complex. The authors developed a taxonomy which mapped the identified relationship between emotion regulation processes and specific emotion regulation strategies and strategy subtypes and further detected several moderators of the effectiveness of emotion regulation strategies containing 15 distinctions. Our study seems to support that corresponding processes are multifaceted. As mentioned above, a general assumption in cognitive psychology is that emotional resources are scarce and that individuals will try to budget in terms of these resources to maintain overall job performance (Joseph & Newman, 2010). An analogous strategy for IR emotion management in the context of virtual world use could thus be to learn which level of IR emotion management is absolutely necessary for performing a certain task and finding out how not to unnecessarily "drain" resources (Joseph & Newman, 2010, p. 57) in order to avoid related distraction. Consequently, IR emotion management capacities will play a more important role for tasks in virtual worlds that require more emotional attention, yet regulating one's own IR emotions may play a secondary role in comparison with acting on others' IR emotions due to the fact that the former does not provide valuable advantages with regard to one's performance for the present task.

The results concerning SE cognitive ability with regard to emotional capabilities seem to be in line with related Joseph and Newman (2010)'s findings, while those with regard to IR performance were rather unexpected; general cognitive ability is assumed be a good predictor of a broad range of achievement indicators like academic performance, career potential, creativity, and job performance (Kuncel et al., 2004). At first, we investigated whether the accuracy of self-reported estimates, as suggested by some (cf. Ackerman & Wolman, 2007; Greven et al., 2009), needed to be questioned; we reasoned that the estimation of cognitive ability could have been biased due to error in participants' perceptions of their performance (Ehrlinger & Dunning, 2003). However, this view would make it difficult to explain why SE cognitive ability had a similar effect on emotional capabilities in our study as in the study conducted by Joseph and Newman (2010). Our conclusion is that in order to relate cognitive ability to IR performance, a more nuanced performance measure than overall performance is needed to better understand the requirements imposed by different virtual world tasks. Attempts to bring together differing understandings of success measures across different fields, for example of the view of educational literature and cognitive psychology on college success (Robbins et al., 2004), point towards this direction. At the same time, institutional achievement (at school, university, etc.), which has shown to be linked to IR emotion understanding in our study, seldom accounts for spatial, navigational, or strategic abilities often required in virtual worlds (Stanney et al., 1998; Sacau et al., 2008). Applied to IR performance, this suggests that institutional achievement feedback may actually account for emotional capabilities, but may be of little value with regard to typical virtual world tasks.

Our findings in terms of cognitive absorption indicate that being cognitively absorbed particularly enhances the emotion-related process of IR emotion perception, and to a lesser extent, IR emotion understanding. This is an interesting finding because of the true IR nature of this relationship, integrating the absorption by an IS / medium like a virtual world to cognitive processes related to emotions. Our findings yet also show that while cognitive absorption may be a good predictor of IS usage, it does not seem to qualify as a predictor of IS performance.

5.2 POSSIBLE LIMITATIONS

To be invited to take part in the biannual survey of the portal operator was an invaluable opportunity which enabled us to conduct a study with a great many of participants (cf. Yee, 2006c). However, as with every research project constrained in terms of budget, time, and other resources, our study is not without its limitations. The *Publication manual of the American Psychological Association* (2011, pp. 35–36) gives a detailed list of suggestions what aspects of a study could or should be dealt with at this point, and we attempt to meet these requirements to the greatest extent possible. The following paragraphs try to identify all potential limitations, sources of bias, and threats to internal validity, and explain how we addressed each of these issues.

LENGTH OF THE SURVEY AND LOTTERY The length of our survey was mainly extended by the additional items of the operator. On that note, many reasons account for the operator's interest in retrieving as much survey data as possible could be enumerated, out of which the most important three are likely to be the following:

1. Sponsoring is an essential pillar of the portal operator's business model. In return for their financial support, sponsors are granted certain privileges, for instance they may determine questions of particular interest to them which survey participants will be asked in the survey that the operator conducts regularly. The number of questions granted is usually in proportion to amount of the payment.

2. Another important source of revenue are advertisement orders, and information collected via the regular survey reveal important details on what products to best advertise on the portal and how.

3. The survey results are important means for convincing new potential business partners to engage in advertising or sponsoring, as the results of the survey reveal valuable information (like purchasing habits etc.) on their target group.

However, in order to safeguard against inconsistent data, we needed to implement many control variables, a fact which may ultimately have lead to a larger dropout rate than might have occurred with a shorter questionnaire (cf. McKnight et al., 2007, p. 68).

By the setting of present study, a scientific research project and a commercial project were deeply interwoven, and effects of this interwovenness on response rate and quality are difficult to account for (just as it is difficult to account for reasons for missing values). Questions could therefore be raised regarding the general acceptance, the motivation to participate in, or, respectively, the reasons for dropping out of the survey at different stages (Göritz, 2004). As Hinkin (1998) points out, "keeping a measure short is an effective means of minimizing response biases caused by boredom and fatigue" (p. 109). As the conduction of such large surveys is costly and their acceptance by the community limited to a certain extent, the operator only launches them every two years. To motivate users to participate and, moreover, to overcome survey fatigue and thereby dropout, the operator makes use of gift lotteries, a method which is popular in marketing and also amongst researchers (Göritz, 2004). For our study, participation had not only been encouraged by (a) raffling prizes, but also by an (b) introductory text displayed on the welcome screen which stated that filling out our questionnaire contributed to studying the phenomenon of eSports in more depth, as well as by the (c) announcement that answering the questions items could help improve the leagues and the personal experience when visiting the portal.

Studies in higher education, health, and marketing have recently compared effects of different types of surveys (web-based or postal, commercial or nonprofit), settings (self-selected or not self-selected), and types of incentives (lottery-based: high or low payout; high or low probability of payout; fixed: conditional or unconditional; material, nonmaterial or no incentives) on survey outcome (Göritz, 2004; Laguilles, Williams, & Saunders, 2011; Halpern et al., 2011; S. R. Porter & Whitcomb, 2003). Results of the aforementioned studies were contradictory to each other when concluding whether incentives had a significant impact on response rates for different settings.

However, Göritz (2004) found that in web-based surveys, whether participants were self-selected or not, there was no difference in their susceptibility to the incentive conditions. Neither in commercial nor in nonprofit settings was response quality or survey outcome affected by different types of incentives, the type of prize yet seemed to have a mild effect on the sample *composition*. The authors reasoned that prizes may be differently attractive to participants depending on age, education, intensity of Internet use, and gender. In our study, the latter two characteristics are likely to be negligible, firstly due to a mostly male and technophile target group, and secondly because the age and education profiles of our survey candidates were quite similar to those found in the survey data of earlier years (with different prizes), indicating that these sample characteristics should not be of concern either.

At this point it may be interesting to note that–as far as we were able to judge from the free comments that were left at the end of the survey–participants seemed intrinsically motivated to answer our questionnaires. Some participants even made suggestions how to improve the questionnaire, while others mentioned that they would like to know about the results of our study. In addition to the abovementioned, this may further indicate that our study has not been biased by a relatively long survey or by lottery incentives to a greater or lesser degree than any other study based on surveys.

SAMPLING PROCEDURE Though the motives of the operator for adding items to our questionnaire are easily understandable, one issue related to this may be the division of the survey into different questionnaires due to an otherwise excessive length of the overall survey and thus the creation of subsamples. In an ideal scenario, our survey would have been sent to all members and not only to every sixth member of the community.

Nevertheless, possible issues as a consequence of this type of sampling were carefully addressed. A potential selection bias was countered by sending our questionnaire specifically to every sixth player ID, which itself is taken from a consecutive series of numbers. The player ID only reflects the order in which users have joined the community and thereby roughly indicates at what date this has taken place; other characteristics are completely unrelated to the ID. Also, IDs of banned players or deleted accounts are not assigned again, which means that the original joining order is always preserved. The algorithm applied ensured that player IDs from all dates of accession since the opening of the portal were evenly distributed among the different questionnaires. When comparing our sample to those of the other questionnaires, we found that all six samples almost had the same size. In addition, the existence of several different questionnaires was revealed to all survey candidates right at the beginning, and this was already a well-known mode for the–supposedly many–players who had taken part in the operator's survey before. We have therefore no reason to believe that the implementation of the survey should contain a systematic error by construction in that respect.

ELIMINATION OF SUSPENDED PARTICIPANTS One alternative to eliminating the suspended profiles would have been to check if the status of some player accounts had changed and to re-crawl the particular profile pages again. This solution was rejected due the fact that too much time would have passed in between the two crawls, which in turn may have lead to measurement distortion. We believe that in order to retrieve comparable and unbiased results regarding performance, collecting all data during the same short period of time was the most reliable alternative, especially with regard to external factors such as public opinion on gaming which could have an influence on social desirability (cf. D. L. Phillips & Clancy, 1972) and similar relevant aspects of survey outcome. Another option was to impute or estimate values for the player level variable by applying a MDT, yet this practice has been rejected by some (cf. Graham, 2003; Allison, 2009).

SELECTION OF CASES FROM NONUNIQUE GROUP During the preliminary data investigation, we discovered that we had to deal with a nonunique group of respondents, that is, a number of survey candidates who appeared more than once in the data. A similar case has, to the best of our knowledge, never been published before. In our view, the fact that we had several answers of the same participant to the same questionnaire at our disposal was a unique opportunity to test for consistency of answers and thereby to examine the reliability of our instrument, unprecedented in literature. The challenge was to find a practical and at the same time expedient way to investigate these entries and to detect unusual pattern in such a large data file with very complex pattern.

Some initial ideas on how to achieve pattern detection needed revision during the course of the investigation. For example, different than expected, the time stamp of entries did not provide any useful information in this regard; whether an entry belonged to the first, the second, or third response "round" of a survey candidate was independent of the total amount of missing values for that particular entry. Another parameter we had considered for comparison and evaluation was the amount of time a participant had needed to fill out the questionnaire (i. e., between two rounds of a participant, but also between participants). According to the operator's staff, however, abilities of members of the target group with regard to eye-hand coordination may range from average to extraordinary, so that analyzing the time spent to complete the questionnaire would have lead to highly speculative results. The approach we finally applied for case selection is explained in depth in Section 4.1.

COLLECTION OF PERFORMANCE DATA In our view, there were three alternative ways to collect the player level scores:

- One option consisted of explicitly asking the survey candidate for this information. This approach was rejected for two reasons: On the one hand, retrieval by questionnaire item could not exclude missing values; on the other hand, retrieval by this means could not also exclude the possibility of bias due to self-reporting.

- Another option was to run specific queries on the portal operator's data base. This turned out to be an impracticable approach due to the amount of explicit queries needed, problems related to checking and double-checking the results, and the difficulties to store the scores in an adequate format.

- The last option (and the one we preferred) was to combine the benefits of unbiased and complete data with feasibility by using a web crawler which would gather performance data from the profile pages via Internet.

Collecting performance data from the league system not only allowed for retrieval of scores for all users, but also allowed for measuring the scores objectively: Player levels are calculated automatically each time a game has ended and cannot be tampered. Consequently, this option provided a solution to two important problems, namely missing values and self-report bias. External manipulation of game scores, for example, by attacking the server, is controlled through extensive intrusion detection. On the basis of these considerations, we believe that our approach to the collection of performance data should not cause any distortion.

Unfortunately, adding our IP to the portal server's whitelist in order to be excluded from the intrusion detection turned therefore out to be too much of an effort in terms of organization and security. Hence, to download the profile pages safely first and to parse them later (i. e., offline) was chosen over parsing them directly (i. e., online), also for the following reasons:

First, and most importantly, this allowed for a fast and flexible correction of parsing errors and thus full control over unexpected results. Most of the errors in the parser code were found through testing. Still, some errors could only be detected when analyzing the data in its entirety, and it turned out that those errors had not been detected through testing because of specific characteristics of the subsample of data we had (randomly, so we thought) selected for testing purposes. However, since the profiles were stored as local files, their re-parsing after the rectification of a parser error only took hours, in contrast to days if data had had to be collected afresh via web crawler first.

Second, this solution featured independence from Internet connections and server availability to a large degree. As we knew from earlier observations, the analyzed Internet portal often struggled with server down times. These down times could sometimes last for days and their occurrence was unpredictable. It was therefore more convenient to download all the profiles quickly one after the other–that is, between two down times–and to check for the completeness of all their corresponding files straight away, rather than to determine which specific data entries were missing or erroneous (e. g., *because* of down times) and to restart the process for those specific entries again. This way we were able to capture the current state of the community under investigation with a time gap of only nine days between the beginning and the end of the crawling process.

Third, the resulting "snapshot" of the community would be preserved for further analysis–as for instance related to our secondary hypotheses–to be tested in the future.

REPRESENTATIVENESS OF THE SAMPLE A large number of different types of virtual worlds were represented in our study, from ego-shooters to social games. Also, in contrast to studies that use university students as research subjects (see, e. g., Compeau, Marcolin, Kelley, & Higgins, 2012, for a criticism of this pratice), we were able to survey our "true" target group and to test our hypotheses in a realistic environment with regard to our research goals. However, we cannot exclude the possibility that the investigated community "may differ in terms of size, maturity, and culture. Future research should endeavor to include various types of virtual worlds so that the results obtained can be more comprehensive and broadly applicable" (Animesh et al., 2011, p. 806).

Furthermore, one has to be aware of the fact that when conducting an inquiry based on sample statistics, one collects sample points that are generalizable only to sample estimates (A. S. Lee & Baskerville, 2003). In our case, the female gender was downright excluded from our study. This was not intended by design nor was it the result of sampling bias, this situation simply occurred because women were disproportionately underrepresented in the investigated population (see Section 3.1). The following aspects support this observation:

- To countercheck the information we had gathered on gender distributions, we compared our data with data from the operator, because the operator not only asks members to indicate their gender when registering to the portal community, but also performs identity checks on the basis of official ID cards (like passports, etc.) when granting certain "trusted player" levels (see again Section 3.1). Thus gender information provided by the operator should be reasonably valid.

- Our assumption that women are by far outnumbered in the community under inspection was also confirmed by the data from other questionnaires (i. e., no. 1–5, see Section 3.4.1.1) of the year of our survey as well as of surveys of earlier years, too.

We are therefore confident that the gender composition found in our sample was close to the "true" composition of the population under investigation and that it is unlikely that a systematic gender bias had occurred (for a word of caution on representativeness and significance, see Cohen, 1994).

Members of the community are aware of the community's gender composition, but they may not necessarily have born in mind that they were being inquired by female scientists. There is also evidence (e. g., from the general context, from the anonymous setting, and from free text comments) that the participants felt somewhat free to express how they felt in a safe environment. To this end, having investigated an almost solely male sample of participants–of a target group that is subject to prejudices in this context anyway–on their emotional skills did not only contain some essential irony, but was also seen as an incredible opportunity.[1]

5.3 MISSING DATA, DISTRIBUTIONS, AND OTHER ISSUES RELATED TO CONDUCTING EMPIRICAL STUDIES

As aforementioned, many of the challenges we encountered during the course of the study were connected to the characteristics of our data. However, topics concerning sample size, normality of the data, model estimation, and so on (cf. Goodhue et al., 2012; Ringle et al., 2012), are common threats to all empirical studies across different fields. Nonetheless, in order to address them adequately, we needed to put a lot of effort into developing proper solutions. In the following, we link prominent data issues to the current literature, evaluate their possible effects on empirical studies, and explain how we solved the respective issues, thereby sharing useful insights into our approach. We first discuss the data-specific issues we encountered and later proceed to more general issues related to the different methods of analysis we used.

Data-Specific Issues

This section motivates in more detail some of our choices of approaches and methods with regard to measurement, data screening and cleaning, data distribution analysis, sample size and power analysis–with a particular focus on the role of missing values.

Missing Values

Missing observations "are the rule rather than the exception in marketing data" (Kamakura & Wedel, 2000, p. 490), are frequently encountered by researchers in psychology (Sinharay et al., 2001), in the medical (D. B. Rubin, 1996), the social, and the behavioral sciences (Schafer & Olsen, 1998; Raykov, 2012) as well as in statistics, economics, biometrics (Roth, 1994), ecology and evolution (Nakagawa & Freckleton, 2008), have been called "unavoidable" in clinical research (Sterne et al., 2009, p. 157), and are a common issue with multivariate data (Paddock, 2002) and particularly with surveys (Enders, 2001b; Andridge & Little, 2010; Grittner et al., 2011; Roth, 1994).

Analyzing missing values and treating them adequately is crucial because their inappropriate handling "can lead to bias in parameter estimates (. . .), bias in standard errors and test statistics (. . .), and inefficient use of the data" (Allison, 2003, p. 545). Inadequate MDTs

1 Interestingly, the editorial board of the *MIS Quarterly* special issue on virtual worlds was an all-women team (cf. Wasko et al., 2011). One could raise the question whether this is related to a self-selection effect which has its roots in the perception of virtual worlds as a social phenomenon, thereby perpetuating the stereotype of women predominantly being interested in topics connected to the social sciences.

also pose a threat to Type I and Type II error rates and confidence interval performance (Collins et al., 2001). Missing values "can affect the quality (i.e., reliability and validity) of our systematic observations[,] .... [the] strength of the study design[, and] (...) the validity of our conclusions about relationships between variables ....[, and can furthermore] limit the representativeness of the study sample" (McKnight et al., 2007, pp. 18–19). In other words, missing values affect construct validity, internal validity, and causal generalization (McKnight et al., 2007). Consequently, APA's task force on the subject recommends to report missing values, their handling, and their potential effects for any type of empirical study (cf. McKnight et al., 2007). Other sources on the subject suggests that reports should include detailed information on how researchers dealt with missing data for each individual variable, what MDT they applied, and for what reasons, in order to "help others understand their analyses and promote replicability" (Roth, 1994, p. 556). However, as researcher often avoid to report how they handled their missing data (e. g., in communication research, cf. Myers, 2011; or in psychology, cf. Roth, 1994; McKnight et al., 2007), it must be assumed that the large majority of researchers still use listwise deletion, the "worst possible of all methods" (Myers, 2011, p. 298; cf. also Enders, 2001a), when facing missing values.

According to our findings, the missing mechanism (also called distribution of missingness, probability of missingness, or (non)response mechanism, cf. e. g., Schafer & Graham, 2002, Sinharay et al., 2001, Andridge & Little, 2010) and the amount of missings are those characteristics which–in relation to the relative impact of missings on study results–usually receive the most attention in literature. Rather scant attention is paid to the patterns of missings and their possible implications (cf. e. g., Allison, 2003; Schafer & Graham, 2002), and the same is true of types of missings (cf. e. g., Schafer & Graham, 2002; McKnight et al., 2007; Graham, 2009). We discuss details on each aspect below.

MISSING MECHANISMS One of the first to investigate the *process* that causes missing data was D. B. Rubin: In his seminal work (1976), he identified the conditions under which this process becomes relevant. Back then he formulated the "statistical relationships between the data and the missingness" (Schafer & Graham, 2002, p. 151) for which missings can be considered ignorable or nonignorable. Later, these observations were supplemented by the joint work of Little and Rubin (1987[2], as cited by Byrne, 2010, p. 354). Researchers today typically classify missings by distinguishing between (a) MCAR, (b) MAR, and (c) MNAR (cf. e. g., Graham & Coffman, 2012; McKnight et al., 2007), though some authors prefer to use the original distinctions made by D. B. Rubin (1976) and hereby speak of MAR, observed at random (OAR), and

2 Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley (as from the references of Byrne, 2010, p. 377).

parameter distinctness (PD; cf. Collins et al., 2001), with the special case of MCAR being a combination of MAR and OAR (Heitjan & Basu, 1996). Though possible reasons for missings may be insightful and of interest to a study (Allison, 2003; van den Broeck et al., 2005; van Ginkel et al., 2007; McKnight et al., 2007; Zhang, 2005), causal relationships–or rather possible or actual reasons *why* missings occur during a specific study–are disregarded with this approach (Schafer & Graham, 2002). The terms MCAR, MAR, and MNAR refer to a statistical point of view entirely (Allison, 2003). Some have argued in favor of an approach "which is meant to use simple English" and prefer to actually address causes of missingness (Graham & Coffman, 2012, p. 277), also because of the fact that the aforementioned terms are potentially misleading due to the perception of "random" in real-life applications (Schafer & Graham, 2002; Graham, 2009; Enders, 2001b). Others argue that the exclusion of reasons from analysis constitutes a major advantage because, as causes may be numerous (Allison, 2003) and complex (cf. McKnight et al., 2007), accounting for all of them is "not realistic" (Schafer & Graham, 2002, p. 150).

Generally speaking, MNAR is considered nonignorable (van Ginkel et al., 2007), ignorable thereby referring to "whether the mechanism of missing data must be modeled as part of the parameter estimation process" (McKnight et al., 2007, p. 50). Conversely, MAR and MCAR are "ignorability conditions–when they hold, they guarantee that certain kinds of inferences may be made without recourse to complicated missing-data modeling" (Heitjan & Basu, 1996, p. 207). However, due to the fact that assumptions required to hold for MAR are untenable (Graham, 2009), no method can be applied to distinguish between cases of MAR and cases of MNAR (Sinharay et al., 2001; Sterne et al., 2009), that is, unless counterchecked through a follow-up data collection (or implied by a model which cannot be tested), it is impossible to test whether MAR holds (Schafer & Graham, 2002). In this context, it is important to comprehend that

> missing data mechanisms are not characteristics of an entire data set, but they are assumptions that apply to specific analyses. Consequently, the same data set may produce analyses that are MCAR, MAR, or MNAR depending on which variables are included in the analysis (Baraldi & Enders, 2010, p. 8).

Some argue that missings in real world data is unlikely to violate nonignorability assumption (Allison, 2003). Others argue that MCAR, MAR, and MNAR may coexist in a data set and that an actual data situation is unlikely to be classified accurately by a single mechanisms alone (Schafer & Graham, 2002; McKnight et al., 2007). Graham (2009) even expressed the view that "the best way to think of all missing data is as a continuum between MAR and MNAR" and that "whether it is

MNAR or not should never be the issue" (p. 567). It is the responsibility of the researcher to evaluate whether the likelihood of a MNAR situation should be of serious concern (Sterne et al., 2009).

Some MDTs exist that do not posit an ignorable missing data situation, namely selection-models and pattern-mixture models, yet with them similar problems as with MAR tests occur: The former requires to (a) make untestable assumptions with regard to the population distribution at a certain stage and are sensitive to minor changes in the assumed shape of this distribution, while the latter require the (b) estimation of model parameters which can neither be supported nor contradicted on the basis of the observed data (Schafer & Graham, 2002). To address the issue of untestable distribution assumptions, (Asparouhov & Muthén, 2009) have demonstrated how to perform a sensitivity analysis by comparing several different models with Mplus. In the future, we might also see more approaches attempting to enhance the plausibility of the MAR assumption like the one proposed by Raykov (2012), which generates point and interval estimates of construct correlations in the presence of missing data, again supported by Mplus.

AMOUNT OF MISSINGS There seems to be no generally accepted definition of what constitutes a large or small amount of missings (sometimes also referred to as the *level* of missings, cf. Mackelprang, 1970) (McKnight et al., 2007), nor do existing guidelines clearly specify benchmarks for the amount of missings in a data set does not require treatment.

When searching for clarification, we experienced that

> amount is a concept used ambiguously in the missing data literature. Most often it refers to the number of subjects for whom data are missing, but it can also refer to the total number of missing observations from a particular variable or set of variables or to the total number of missing observations from a data set (McKnight et al., 2007, p. 61).

For example, articles reporting studies that have tested different MDT for different missing percentages often fail to state whether percentage specifications apply to a single variable or to a whole data set,[3] and authors tend to use nonspecific quantifications like "small to moderate amounts of missing data" (e. g., Allison, 2003, p. 551). Or, even if quantifications are precise, what they refer may not become clear: Roth (1994) stated for instance that "Monte Carlo studies suggest there is little difference in the parameter estimates and answers to research questions when less than 10% of the data are missing" (p. 550, that is, for

3 In addition to that, variable distributions and missing patterns underlying the data used for the tests are usually not specified, see Collins et al. (2001) for an example. The opposite situation is not rare either: Enders (2001b), for example, examines the impact of nonnormality on the performance of various MDTs, yet remains vague in terms of amount.

both "random patterns (. . .) or systematic patterns", q. v.), and Kline proposed that missings "should probably constitute less than 10% of the data" (1998,[4] p. 75, as cited by Byrne, 2010, pp. 353–354) and that a "large" (p. 75, as cited by Weiber & Mühlhaus, 2010, p. 144) amount of missings started at 10%. Historically speaking, calls for the treatment of even small fractions of missings seem to have grown louder as computational power and the number of tools available have increased;[5] more recent publications have quantified "a small fraction" of missings as 1–4% (e. g., Andridge & Little, 2010, p. 41; Graham, 2009, p. 554).

PATTERNS OF MISSINGS Roth (1994) has only distinguished between two types of pattern, random and systematic, whereas Schafer and Graham (2002) have used more distinctions and categorized patterns as either (a) univariate, (b) monotone, or (c) arbitrary (see Figure 8); according to the latter, the first pattern occurs when there are only missings on one single variable or, alternatively, when one can identify blocks of items which, for each participant, are either completely observed or completely missing; the second pattern describes a condition where items (or respectively, groups of items) can be ordered in a way that if one item (or group of items) is (are) missing for a participant, all subsequent items (or groups of items) will be missing as well; the third pattern is present when, at least from appearance, no regularity can be detected. Paddock (2002) and Allison (2003) both agree that real-life data sets will typically exhibit arbitrary–sometimes numerous and overlapping–patterns of missings.

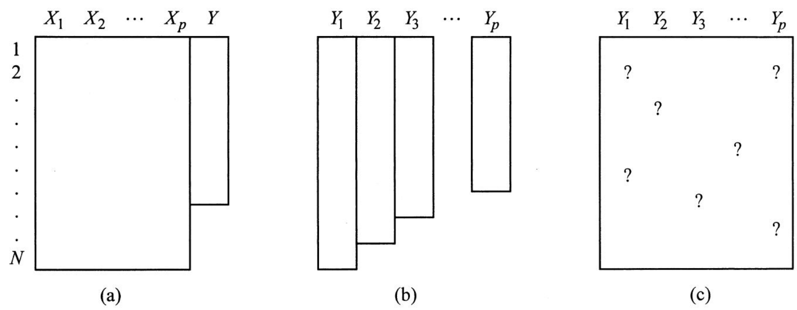


Figure 8: Patterns of nonresponse in rectangular data sets: (a) univariate pattern, (b) monotone pattern, and (c) arbitrary pattern. In each case, rows correspond to observational units and columns correspond to variables. Adapted from: Schafer, J. L., & Graham, J. W. (2002). Missing data: Our View of the State of the Art. Psychological Methods, 7(2), 150.

---

4 Kline, R.B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford (as from the references of Byrne, 2010, p. 376).

5 For example, we meanwhile found out that in a following edition of the book mentioned earlier, Kline formulated a lot more carefully and precisely in saying that "a few missing values, such as less than 5% on a single variable, in a large sample may be of little concern" (2011, p. 55).

Pattern identification is an integral part of pattern-mixture models (see above) or the multi-group approach, both procedures that are intended to address missings; for this purpose, participants are sorted depending on their patterns of missings, and the emerging clusters are expected to provide clues with regard to subsets or groups (Schafer & Graham, 2002). In a study with n types of measurement as many as 2^n different missingness patterns, so "manual calculation can be extremely laborious" (Campbell et al., 2007, p. 409). We too considered to apply the aforementioned procedures, yet when examining our data, we did not discover any such clusters. A graphical representation of the patterns observed in our data is displayed in Appendix C, Section C.2.

The few fruitful comments in terms of implications of missing patterns for MDT we were able to find were made by Allison (2003), who suggested that "hot deck and predicted mean matching methods can work well (...) when the missing data follow a simple 'monotonic' structure" (p. 555; note that no missing mechanism or amount of missings was mentioned in the context of that suggestion), and by Zhang (2005), who stated that "the propensity score method (...) and the predictive model method"–two types of MDTs based on multiple imputation–"are for the missing data with a monotone pattern" (p. 142).

TYPES OF MISSINGS The traditional type classification of missings distinguishes between (a) unit nonresponse, (b) item nonresponse, and, in the context of longitudinal studies, (c) wave nonresponse. With this taxonomy, attrition, or dropout is considered a special case of wave nonresponse (Schafer & Graham, 2002). In this work, we have used the term dropout to cover a situation which "occurs when one leaves the study and does not return" (see Schafer & Graham, 2002, p. 150). In our study, once a participant had dropped out at a certain point, because the implementation of our survey system was not designed for stable sessions, the decision was irreversible, hence a participant could not resume the survey once he or she had left the system. Dropout is considered the type of nonresponse which is the most common (see Schafer & Graham, 2002; Graham, 2009; McKnight et al., 2007), and "virtually ubiquitous" (Graham, 2009, p. 567) in longitudinal studies. Graham (2009) proposed that "the effects of attrition on study conclusions in a general sense are not nearly as severe as commonly feared" (p. 567).

Not surprisingly, we discovered that types and patterns of missings were apparently not mutually independent; for obvious reasons, dropout due to survey discontinuation lead to distinctive monotonic pattern (cf. also McKnight et al., 2007, p. 62).

As stated previously, missing values are prevalent in many studies, and the average amount of missings exceeds 30% by far–up to 50%, according to results from psychology and marketing, for instance (McKnight et al., 2007, p. 3; Kamakura & Wedel, 2000, p. 491). However common the situation, a researcher encountering missing data is potentially faced with serious difficulties: As Mackelprang (1970) put it in the context of a study on pairwise deletion, "missing data constitute either measurement error or sampling error, depending on how the missing data are handled" (p. 495), and the higher the rate of missings, the higher the suspected impact on parameter estimates and on explanatory power of conclusions (McKnight et al., 2007). Also, as mentioned in Section 4.1, many MDTs with desirable characteristics have only been tested on (simulated) data with an absolute maximum of 50% missings (cf. e. g., Black et al., 2011, p. 1845; Yuan et al., 2004, p. 422; and Graham, 2009, p. 560). On top of that, these MDTs make assumptions which will typically not hold with real data, like, for example, multivariate normal distributions for ML-based MDTs, or make assumptions which are untestable, like MAR (cf. Sinharay et al., 2001; Enders, 2001a; Schafer & Graham, 2002; Allison, 2003).[6] Yet when simply deleting cases with missings, a researcher loses a vast percentage of the total data set (Roth, 1994), thereby obtaining substantially biased estimates (Enders, 2001b; Raykov, 2012) and reduced statistical power (Nakagawa & Freckleton, 2008).

Missings are also a nuisance for practical reasons: As Allison (2003) expressed it, "virtually all methods of statistical analysis are plagued by (. . .) missing data" (p. 545), because, from a technical point of view, "most data analysis procedures were not designed for them" (Schafer & Graham, 2002); therefore, if missings are not treated explicitly, statistics software will either not run (like AMOS) or apply a MDT implicitly (like SPSS). With regard to the latter, listwise deletion is the "default in virtually all statistical packages . . . . Because this produces a working sample with no missing data, any statistical method may then be applied" (Allison, 2003, p. 547; see also Enders, 2003). When performing "internal consistency analyses with item-level missing data" (Enders, 2003, p. 322) with SPSS, for example, the program will automatically implement listwise deletion, without prior or subsequent warning. However, using this kind of ad hoc procedure is generally only safe under very strict conditions (namely when the MCAR assumption holds) and may otherwise yield biased parameter estimates, loss of power, and other negative effects (Bühner, 2011; Allison, 2003; Graham, 2009; Sinharay et al., 2001; Finch et al., 1997). Hence Savalei and Bentler (2009) have come to the conclusion that "in SEM, the old ad-hoc approaches, such as listwise and pairwise deletion, hot deck imputation, and so on, are no longer deemed acceptable" (p. 477).

On the bright side, "methods that assume ignorability would still perform very well if the data are merely MAR" (Allison, 2003, p. 545). In a study reported by Enders (2001b), CFA and SEM estimates obtained

6 Note that Allison (2003) claimed that "ML for missing data can be implemented under a variety of distributional assumptions" (p. 548), yet gave no further details on the subject and only discussed methods "that are based on multivariate normality" in the following (q. v.).

under a MAR mechanism were unbiased (different from the estimates obtained when cases with missings were removed from analysis), and the bias introduced to ML estimates by MNAR, that is, a nonignorable missing mechanism, was still less than the bias observed for listwise and pairwise deletion. In another study on MI, bias introduced by the missing mechanism showed to be comparatively small (van Ginkel et al., 2007). Moreover, unlike in ecology and evolution studies in which organisms may die prematurely, thus *before* being able to express a particular trait of interest (Nakagawa & Freckleton, 2008), or in clinical studies “in which the reasons for dropout may be closely related to the outcomes being measured” (Schafer & Graham, 2002, p. 172–173; cf. also Grittner et al., 2011), so different to the aforementioned situations in which MNAR simply has to be presumed, it can be assumed that for studies conducted in the social sciences, even if there was a relation between dropout and outcome, such a relation would only introduce minor bias, and the application of a MAR-based methods would still be acceptable (Schafer & Graham, 2002). Also, the performance of many advanced MDTs is still very good at the 50% missing level (e. g., Black et al., 2011), so that, in our view, levels slightly above that threshold–like those found in our data–should only have a minor additional impact on MDT outcomes. And finally, it has been suggested that “loss of power due to attrition has frequently been overestimated because the impact of modern missing data procedures on power has not been considered” (Graham et al., 2012, p. 254).

To summarize: The importance of the missing mechanism has been affirmed by a large number of authors, and by many, this aspect is considered more important than amount, pattern, or type of missings. In fact, McKnight et al. (2007) suggested that for parameter estimation, a situation with a large amount of missing observations are which ignorable might be favorable to a situation with only smaller amounts of missing observations that are nonignorable. And most importantly, MDTs which assume MAR are relatively robust, even if a departure from this assumption has to be suspected (cf. Allison, 2003; Schafer & Graham, 2002).

Choice of Missing Data Treatment

Deletion techniques are “exceedingly common in disciplines such as psychology and education” (Baraldi & Enders, 2010, p. 6), and recommendations when and under what circumstances not only to use, but to actually favor them can be found in literature across fields. For instance, Mackelprang (1970) clearly advocated to use pairwise deletion before applying factor analysis;[7] in line with this recommendation, Roth (1994) cautioned that imputation techniques “will artificially increase clarity of factor structures” (p. 556) when factor analyzing data,

7 To qualify this statement, this was said at a time when newer techniques were still “in experimental stage” (Mackelprang, 1970, p. 505).

whereas the use of listwise or pairwise deletion would prevent this type of distortion.[8] Graham (2012a) has reasoned that "complete cases analysis tends to perform quite well" (p. 48) for analysis of covariance or multiple regression analysis, under the condition that several predictors from a pretest are available and that only a single dependent variable is implemented.

However, parameter estimates from data sets which have been trimmed be deletion techniques are often incorrect when data are not MCAR (Nakagawa & Freckleton, 2008). The report of the APA Task Force on Statistical Inference (1999)[9] even concluded that MDTs which consist of deleting cases "are among the worst methods available" (p. 598, as cited by Baraldi & Enders, 2010, p. 6). As this view is shared by many, we attempted to identify alternative MDTs which, on the one hand, would suit our data (in terms of distributions, pattern, etc.) but, on the other hand, would not suffer from the deficiencies associated with conventional methods. Though the basic procedures have been available for quite some time (cf. Collins et al., 2001), such high-performance approaches are often referred to as "modern" (cf. Baraldi & Enders, 2010; Schafer & Graham, 2002). Their categorization is often not clear-cut and seemingly contradictory. To differentiate them precisely, one has to pay attention to a whole range of small details (cf. e. g., McKnight et al., 2007; Graham, 2009, 2012a; Zhang, 2005; Sinharay et al., 2001). In an attempt to avoid misleading labels and confusion, we only introduce one very broad MDT categorization–out of many possible–and rather focus on technique properties with specific regard to our particular study.

The most widely recommended (or even praised) types of "modern", that is to say, state-of-the-art MDTs (see, e. g., Schafer & Graham, 2002; Graham & Coffman, 2012; Enders, 2003; McKnight et al., 2007; Roth et al., 1999; Baraldi & Enders, 2010) are doubtlessly, on the one hand side

MI estimation techniques (cf. D. B. Rubin, 1996) based on a

A. Bayesian (Schafer & Olsen, 1998; Schafer & Graham, 2002), Monte Carlo (Schafer & Olsen, 1998; Schafer, 1999), and/or[10] "Bayesian-like approach (...) of MCMC" (McKnight et al., 2007, p. 172), respectively, or[11] on an

8 Despite the fact that the author was aware of modern MDT and discussed ML imputation and the EM approach in the same article.

9 Wilkinson, L. American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594–604 (as from the references of Baraldi & Enders, 2010, p. 37).

10 As can be seen throughout the upcoming section, MI terminology is not coherent in many different ways.

11 In fact, Allison (2012) refers to *the* Bayesian approach to MI and states that there are "two major iterative methods for doing multiple imputation for general missing data

B. algorithm known as "the fully conditional specification (FCS), sequential generalized regression (...), or multiple imputation by chained equations (MICE)" (Allison, 2012, p. 4), and

ML estimation techniques like

A. EM-based procedures (for more information on the EM algorithm, see Dempster, Laird, & Rubin, 1977; Do & Batzoglou, 2008) or

B. FIML (cf. Enders & Bandalos, 2001),[12] also known as "raw" or "direct" ML (Allison, 2003; McKnight et al., 2007)

on the other.[13] The basic idea of MI- and ML-based MDTs is similar. Both types treat missing values as random variables and average over them "as a source of random variation" (Collins et al., 2001, p. 331), only that ML methods average with the aid of numerical means and that MI relies on a Bayesian (D. B. Rubin, 1996) and/or[14] Monte Carlo technique (Schafer, 1999), respectively. Yet there are also more general differences, for example:

- MI is a three-step approach (Orton & Ipsitz, 2001): One first imputes the missing values of a data set multiple times in a row, thereby generating m imputed data sets (e. g., $m = 40$) which slightly differ from each other and for which "imputed values represent random samples from a distribution of plausible replacement values for the missing data" (Baraldi & Enders, 2010, p. 16). The desired analysis is then performed with each of these m complete data sets, and m analysis results obtained this way are stored (Sinharay et al., 2001). Finally, the m results are combined into one result according to Rubin's (1987,[15] as cited by Graham, 2012c, p. 57) rules or formula (Graham & Coffman, 2012; Sinharay et al., 2001), respectively. For more details on the MI procedure, see also Allison (2003) and Schafer and Olsen (1998).

- With EM, in a first step called the "E(stimation)" step, missing values are imputed by predicted scores through consecutive regressions "in which each incomplete variable is regressed on the

patterns: the Markov chain Monte Carlo (MCMC) method and the fully conditional specification (FCS) method" (p. 3).

12 Confusingly enough, some authors use the term FIML to designate an analysis based on maximum likelihood estimation on complete data' (cf. Gignac, 2013)

13 Allison (2009) also mentions inverse probability weighting as a promising third option, which however, has "not yet reached the maturity of the other two methods" (pp. 72–73).

14 Schafer is cited by McKnight et al. (2007) on this topic as follows: "MCMC procedures are loosely allied with Bayesian estimation procedures, but Schafer (1997) argues that MCMC might be somewhat *mislabeled* [emphasis added] as Bayesian" (p. 166).

15 Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley (as from the references of Graham, 2012c, p. 69).

remaining variables for a particular case" (Kline, 2011, p. 59). In a second or the "M(aximization)" step, the "whole imputed data set is submitted for ML estimation" (q. v.). The two EM steps are performed over and over again, until the algorithm converges and a stable solution is found–a means to correct for underestimation of variance which is not unlikely to occur with imputation (Allison, 2003). For more details about the standalone EM algorithm, see Do and Batzoglou (2008).

- In contrast, the FIML algorithm, a MDT of the family of ML procedures, does not replace missing values (McKnight et al., 2007). Instead it partitions the cases in a file into subsets with the same pattern of observations, extracts means, variances, and so on, from each of these subsets, and calculates parameter estimates and their standard errors directly from the available data (Kline, 2011). The algorithm thereby reads in the raw data and maximizes the ML function "one case at a time, using whatever information is available for each case" (Graham, 2012a, p. 53); for more on FIML, the reader is referred to Allison (2003) and Enders and Bandalos (2001).

Note that issues regarding the adherence to certain ethical standards apply equally to the use of MI and FIML (i. e., as one instance of ML) (Graham, 2012b, p. 106).

In general, the more variables are included in the MDT computation, the better (Muteki, MacGregor, & Ueda, 2005; Collins et al., 2001): as a means to make the MAR assumption more plausible (Enders, 2003), to diminish the impact of nonignorable missings (Zhang, 2005), and to reduce effects of dropout on internal validity, external validity, and statistical power of the study (Graham & Collins, 2012). Such a MDT strategy is often called inclusive, as opposed to a restrictive (Collins et al., 2001; cf. also Graham, 2009). As mentioned above, variables which are not necessary for the actual analysis model but may, if included in the computation, improve the performance of a MDT procedure are often referred to as auxiliary variables (Collins et al., 2001; Baraldi & Enders, 2010; Allison, 2003). High or at least moderate correlations of the auxiliary variables with the variables of the analysis model are thereby desirable (Allison, 2012; Baraldi & Enders, 2010) because otherwise, the improvement provided may not be substantial (Graham, 2009). As algorithms may not converge when too many variables are included (e. g., a lot more than 100), guidelines on how many variables to base the estimation on and how to choose an adequate variable subset for a MI imputation model, for example, have been published (e. g., Graham, 2009, 2012e; Graham et al., 2012).

The abovementioned MDTs and related tools have been recommended for different types of analysis procedures. We discuss the most important facts with regard to the study at hand below. A discussion of further (mostly commercial) programs using modern MDTs is also provided by Orton and Ipsitz (2001), Allison (2003), and Savalei and Bentler (2009).

MDT FOR EFA AND RELIABILITY CHECKS Because (a) EM generates "the best" ML estimates for means, standard deviations, and correlation matrices (Graham, 2009, p. 556), because research suggests that (b) "alpha estimates obtained from an EM covariance matrix should be superior to those obtained from traditional MDTs" (Enders, 2003, p. 323), because the analysis of Cronbach's α (or coefficient α, respectively,) or an EFA "do not require the overhead associated with MI" (Graham, 2012a, p. 63), and finally, because it can be assumed that (c) "in any given sample, an EM estimate of coefficient alpha will be a more accurate reflection of the population reliability" (Enders, 2003, p. 335), an EM-based procedure appeared to be the first choice MDT prior to our reliability analysis.

Unfortunately, using the EM implementation provided by SPSS is discouraged: "Good" implementations of EM add a correction factor to each imputed value, that is, a random error term, in order to account for "lost variance" after imputation (cf. Graham & Coffman, 2012), whereas SPSS only "writes data out without adding error .... This is known to produce important biases in the data set" (Graham, 2009, p. 556). Parameter estimates generated in such a way will be true ML estimates, different than estimates from real-life data, which, in the more usual cases, are generally less efficient (Allison, 2003). As there was no evidence indicating that the SPSS version we used provided for correction after EM imputation, we needed to evaluate other options. According to Graham (2009), "good" implementations of EM are provided by different programs, as a side product of MI imputation performed by the SAS/STAT software for instance (or, more precisely, the SAS Proc MI), a commercial program by the SAS Institute (cf. also Graham, 2012d). Another one is also applied by the standalone EM imputation program EMCOV (Graham, 2009), which "generates m imputed data sets using the bootstrap technique" (Enders, 2001a, p. 138); EMCOV, as can be taken from the reference manual,[16] produces an (uncorrected) full data matrix with missing values imputed through EM and an additional data matrix of residuals (which would have to be combined to obtain corrected data). The disadvantage of both aforementioned programs is that they do not actually impute the missing values of the original input data set but generate covariances for all (missing and available) data points instead, and that their output solely consists of an "ML estimate of the population covariance matrix (often dubbed the EM covariance matrix)" (Savalei & Bentler, 2009, p. 478). Such an output is only useful if the procedure to be applied subsequently can make

16 EMCOV Reference Manual, v2.2 and v2.3, September 14, 1993 (updated August 11, 1995)

use of the covariance matrix. NORM, which uses "the Markov Chain Monte Carlo (MCMC) algorithm based on linear regression" (Allison, 2009, pp. 81–82), or as some term it, "Bayesian simulation" (Enders, 2001a, p. 138), is the only noncommercial program we know of that not only provides a covariance matrix as output, but also a complete EM imputed data matrix for which all missing values of the input data have been imputed and corrected. As a complete data set seemed to offer a maximum of flexibility (cf. Schafer & Graham, 2002), we chose to use the EM procedure of the NORM program prior to our factor and reliability analysis.

Following the abovementioned guidelines on how to choose the right variables for imputation, we started with the set of variables that was integrated in our analysis model. We then identified all our control variables as well as those of shared interest with the operator (see Section 3.2), selected those among which had the same measurement of scale (see item 5.3), and included them into the imputation model. The remaining ones had to be excluded, because mixed models are not yet supported by the NORM program. Finally, because it is considered problematic to omit the dependent or outcome variable from the MI imputation procedure (Sterne et al., 2009; Graham, 2009; Allison, 2009) as this would imply that the dependent variable is uncorrelated with the other variables in the imputation model (Graham, 2009), we also added the player level, although the variable itself exhibited no missings.[17] Through this, we were able to include as many variables related to our research questions as possible and thereby maximize the plausibility of MAR, yet ensured the algorithms would converge in a reasonable amount of iterations (cf. Graham, 2012e).

During the further procedure, we also needed to decide which cases to remove from the imputation model. On this account, because keeping them was difficult to justify, we followed the recommendation to exclude all cases which had missings on all variables of the analysis model, although technically, programs like NORM are perfectly capable of imputing values appropriately even under such conditions (Graham, 2012e, pp. 79–80). In our case, this applied to the group of cases which exhibited missings on the variables of the primary constructs (see Appendix C, Section C.2 for a graphical representation of the different patterns found in our data),[18] as the focus of our analysis mainly lay on them. In order to control for selection bias, we analyzed age and player level for the remaining 4,219 cases and the excluded 1,369 cases and compared the distributions of the two variables in both groups.

17 According to Allison (2009), the dependent variable itself should not be imputed, and cases with missings for that variable should be deleted in the absence of auxiliary variables which are sufficiently correlated with the model variables (unless under a MAR condition, of course).

18 Since no single data point was missing for the player level variable, it was not considered for this examination.

MDT FOR HYPOTHESIS TESTING As much as the EM approach has been recommended for reliability analysis (particularly involving Cronbach's α) and EFA (e. g., Graham, 2012b), as much has been argued against EM for the purpose of generating data to be used for hypothesis testing (e. g., Graham, 2012e). Standard errors generated during analysis on the basis of an EM imputed data set, for example, tend be too small–possibly substantially (McKnight et al., 2007; Graham, 2009; Allison, 2003), because the "influence of the missing data is not estimated and therefore cannot be used to correct the standard error estimates" (McKnight et al., 2007, p. 166). It is generally agreed that hypothesis testing–including, for example, CFA–should instead be carried out with MI imputed data or with data preprocessed by a ML procedure like FIML (available in, e. g., AMOS, Arbuckle, 2013), respectively (Graham, 2012d, 2009).

MI is a process which accounts for the variability of the imputations and which generally provides consistent estimates of the parameters and their standard errors (Orton & Ipsitz, 2001). Another important argument in favor of this MDT is that with MI, it is very comfortable to add auxiliary variables to the computation (Savalei & Bentler, 2009; Graham & Coffman, 2012). And finally, MI also bears the advantage that it creates complete data sets which, in theory, should allow for any type of subsequent analysis (Sinharay et al., 2001). Several MI tools are available; according to their documentation, help pages, cited references, and literature, MI procedures in SPSS, Mplus, SAS PROC MI, R, and so on, they are generally similar, but there are differences in their implementations (cf. e. g., Allison, 2012). Randomness in NORM, for example, is generated by drawing random values of parameters with the help of a MCMC algorithm and averaging across the random samples produced by this procedure, which causes–typical for MI (Allison, 2009)–slightly different results to be produced each run (unless one uses the same random number seed, Arbuckle, 2013). In our case, the simplest option was to use our preprocessed EM imputed data as a basis for generating more imputed data sets–hence performing *multiple* imputation–with the help of the NORM program. This way we would have followed the typical NORM procedure, as the latter always performs an EM estimation before running the actual imputation procedure. Because with our data, EM had converged after 116 iterations, we configured NORM in such a way that it ran 116 data augmentation steps using 40 imputed data sets each,[19] summing up to $116 * 40 = 4,640$ data augmentation cycles to be performed. NORM (like AMOS and other programs) provides diagnostics plots to check the convergence at the end of a completed MI procedure (cf. Arbuckle,

19 As recommended by Graham, 2012e, 2009 for data with 50% missing information in order to compensate for the power falloff in comparison to FIML, for example. The rule of thumb of $m = 5$ is not sufficient with more than moderate amounts of missings, see also Allison (2003, 2012).

2013; Graham, 2012e). The program ran without error and produced 40 data sets as expected, with satisfactory results.

The next step was to find a suitable way for conducting a CFA for every single construct on 40 different data sets and subsequently combining or *pooling* the 40 analysis results of each CFA (i. e., according to Rubin's rules). Such computational complexity can only be dealt with by an automated process (Graham & Coffman, 2012) which NORM does not offer, thus we needed to find a tool that did. Our search results are subsumed in the following:

Until today, EQS does not provide any implementation of MI. According to its manual (IBM Software Group, 2013, pp. 18–20), SPSS currently offers several procedures that take care of pooling automatically when provided with the MI module, namely (grouped for clarity):

- FREQUENCIES, DESCRIPTIVES, MEANS;
- ONE-SAMPLE T TEST, INDEPENDENT-SAMPLES T TEST, PAIRED-SAMPLES T TEST;
- ONE-WAY ANOVA, LINEAR MIXED MODELS, GENERALIZED LINEAR MODELS AND GENERALIZED ESTIMATING EQUATIONS;
- BIVARIATE CORRELATIONS, PARTIAL CORRELATIONS;
- LINEAR REGRESSION, BINARY LOGISTIC REGRESSION, MULTINOMIAL LOGISTIC REGRESSION, ORDINAL REGRESSION, COX REGRESSION;
- DISCRIMINANT ANALYSIS, CHI-SQUARE TEST, BINOMIAL TEST, RUNS TEST, ONE-SAMPLE KOLMOGOROV-SMIRNOV TEST, and
- TWO-INDEPENDENT-SAMPLES TESTS, TESTS FOR SEVERAL INDEPENDENT SAMPLES, TWO-RELATED-SAMPLES TESTS, TESTS FOR SEVERAL RELATED SAMPLES.

MI performed by Mplus can be followed by frequentist estimators analysis (cf. e. g., Hjort & Claeskens, 2003) or the estimation of a growth model (cf. Asparouhov & Muthén, 2009); a test of equality of means across latent classes is also provided (with ML estimation), and so are so-called plausible values for latent variables (not factor scores, Mplus Support, 2007). The authors of the NORM program have also developed the MIAutomate[20] tool (yet it needs to be said that it did not run on our computer and needs an instance of SPSS to do the actual analyses). It is supposed to automate the pooling required for multiple regression analysis; for any other type of analysis though, the "combining of results is somewhat less automated, involving some copying

20 http://methodology.psu.edu/db/node/163

and pasting" (Graham, 2012b, p. 95), bearing the risk of introducing serious flaws (Graham, 2012e). The crux of the problem is that

> "multiple imputation requires that the output of one's statistical analysis be a parameter estimate and the corresponding standard error. Multiple regression fits nicely into this requirement in that one always has a regression coefficient (parameter estimate) and a standard error. Other common procedures such as analysis of variance (ANOVA) can be used with multiple imputation, but only when the ANOVA model is recast as the equivalent multiple regression model (Graham, 2012c, p. ix).

To the best of our knowledge, not only is there no example in literature that shows how the required recasting could be done, in fact, no one even seems to have raised the topic in the context of CFA or SEM and argued in favor or against such an approach.

In addition to these practical obstacles, there are also concerns that have been raised with regard to theoretical aspects. After being an widely-known advocate of MI for many years, Allison (cf. e. g., Allison, 2012) has even turned into a MI skeptic.[21] Besides the obvious problems that come with the complexity of applying MI, he particularly criticizes the problems that arise due to the fact that MI uses two models (see also Schafer, 2003), one being the imputation model, the other one being the analysis model. This is problematic because interactions or product terms are nonlinear combinations of two variables, and a regular linear imputation model will not account for them, that is to say, it will assume a correlation of 0 between the interaction term and the dependent variables included in the imputation model (Graham, 2009). To solve this issue, a researcher is required to "anticipate any interaction terms and include the relevant product terms in the imputation model" (p. 561, q. v.). Further important drawbacks of MI are related to the uncertainty arising from the numerous decisions that have to be made during the MI process (Allison, 2012, 2003), or to the extremely differing results for each imputation. Also, if data are collected with a complex sample design and the fraction of missings is large, certain MI estimators may be sizably biased and should be adjusted (J. K. Kim, Brick, Fuller, & Kalton, 2006).

The problems just listed do not arise with ML: It only uses one model (Schafer, 2003) or, rather, estimates the linear model and the probit model simultaneously (Allison, 2009), and "if the model has nonlinearities and interactions, those will automatically be incorporated into the method for handling the missing data" (Allison, 2012, pp. 6–7). Also, the algorithm is deterministic, hence it will produce the same result for a given data set every time it is run (Allison, 2009). However, there are some issues related to the use of ML as well. Performing bootstrap-

21 http://www.statisticalhorizons.com/ml-better-than-mi

ping, for instance to correct for nonnormal data (Byrne, 2010) or to choose among estimation methods (Arbuckle, 2013), is not possible when applying FIML at the same time (Graham & Coffman, 2012; Bühner, 2011). Modification indices, which are useful in case a model needs to be respecified (Chou & Huh, 2012), cannot be provided (though some generally advise against making changes to a model without a compelling theoretical argument Roberts, Thatcher, & Grover, 2010). Also, with the FIML algorithm implemented in AMOS for instance, dealing with missing data requires fitting the saturated and independence models in addition to the actual model (and consequently, the estimation of more parameters), a task which may be quite costly in terms of computation (Byrne, 2010) and may not always complete successfully (Arbuckle, 2013). And finally, though the ML paradigm theoretically allows for the inclusion of auxiliary variables as a strategy to reduce bias and increase power (cf. Baraldi & Enders, 2010), "existing user interfaces and documentation for SEM software do not make it clear how to do this or even raise the possibility that it may be necessary" (Collins et al., 2001, p. 335), thereby practically imposing a restrictive strategy: To provide the option of adding auxiliary variables, a tool is required to enable relatively complicated alterations of the model specification (Graham & Collins, 2012; Savalei & Bentler, 2009), namely ensuring that the model allows "for correlations of each auxiliary variable with (1) all measured exogenous variables and (2) the error terms for each measured endogenous variable" (Allison, 2003, p. 550). Authors have therefore demanded to revise ML software in order to facilitate the handling of auxiliary variables (Collins et al., 2001). Yet until now, only two programs, both of which are commercial, seem to offer a corresponding feature, namely Mplus and EQS (cf. Graham & Collins, 2012; Baraldi & Enders, 2010). Their results, however, need to be treated with caution (Savalei & Bentler, 2009, see also below). As for AMOS, such a feature is not implemented yet.[22] Yet as with EM, in order to minimize bias and to increase power potential, these computations should ideally also include auxiliary variables (e. g., Collins et al., 2001; Baraldi & Enders, 2010). Allison (2003), Graham (2003), and Graham and Coffman (2012) have demonstrated how to adequately a model to perform SEM with FIML and auxiliary variables in AMOS. However, the use of the graphical interface for that particular application is not at all convenient, especially with large models like ours (cf. Graham & Coffman, 2012), as the number of relationships that have to be modeled grows exponentially with every additional variable; hence this approach is quite error-prone. All aforementioned authors therefore preferred to

22 Note that McKnight et al. (2007) seemingly contradict this view in saying that "not all of the model-based procedures make use of the same set of observed data. For example, (...) FIML (...) uses the entire data matrix (...), even when some of the variables in the matrix are not included in the statistical analyses of interest" (p. 172). This could possibly (and mistakenly) be interpreted to mean that all available variables are automatically included in the FIML procedure.

use AMOS' text interface and provided examples of AMOS program code. Unfortunately, the latter approach requires great effort (Graham & Collins, 2012; Savalei & Bentler, 2009; Collins et al., 2001) and is similarly error-prone, as the same amount of correlations is needed (Allison, 2003; Graham & Coffman, 2012);[23] Additionally, as computing statistics like χ^2 and other fit measures needs details on the saturated and the independent model, a user trying to use auxiliary variables is required to supply additional code for fitting these two and explicitly passing these details to the AMOS engine.[24] If these instructions are missing, no fit index can be computed. To our astonishment, we have yet not seen this type of code in any of the examples mentioned in this thesis, nor have we read warnings in this regard.

In our case, since the provided program code examples have been designed for SEM and not CFA, performing a CFA with auxiliary variables would also have demanded to write our own and unapproved code, which in turn had involved the risk of serious flaws remaining undetected. Moreover, because we needed to compare the CFA results of many different models for every single construct, many of such programs would have been necessary. On top of that, it is not sufficient to write a single program that fits all, but a complex program needs to be written for every individual CFA–alternatively, if using the graphical interface, an individual model must be specified each time. Notwithstanding the disadvantages with reference to using auxiliary variables, FIML was thus the MDT of our choice for hypothesis testing, due to its many other advantages (cf. Allison, 2003)–and not least because MI was impracticable for our application.

Considering the statistical benefits to be expected, some have even doubted whether the additional effort is worthwhile and cautioned that altering a model in the required way may lead to “undesirable effects” (Collins et al., 2001, p. 331). In response to this, FIML solutions for SEM have been proposed that are supposed not to alter the substantive aspects of a specified model (Graham, 2003; see also Allison, 2003; Graham & Coffman, 2012). However, these solutions are yet not trivial and, like to the solutions provided by Mplus and EQS, may cause problems due to large model size and underidentification if “the number of A[uxiliary]V[ariable]s is large, or if the residual variances associated with some of the variables are small” (Savalei & Bentler, 2009, p. 478). Also with regard to our own experiences, it is not surprising that a typical FIML study misses out on the opportunity to enhance FIML performance via auxiliary variables and only takes the actual model variables into account (cf. Graham, 2009).

23 We have performed a CFA of CIRME using *age* as an auxiliary variable. The graphical representation can be found in Figure 4. The respective program code can be found in Section E.2.

24 http://www.amosdevelopment.com/support/tips/basic_allfitmeasures.htm.

Normal Distribution Assumption

The fact that many commonly used statistics procedures assume multivariate normality (McDonald & Ho, 2002)–especially those based on ML (Bernstein & Teng, 1989)–keeps being restated like a mantra, and so are the possible consequences of nonobservance. When applying ML covariance structure analysis for instance, test statistics and standard errors are liable to be biased under severe nonnormality (Chou et al., 1991; Yuan et al., 2005; Bentler & Dudgeon, 1996; Sterne et al., 2009). But normality is not only an issue in statistical hypothesis testing, it is also an important factor to consider with regard to the treatment of missing data (Sinharay et al., 2001; Schafer & Graham, 2002; Savalei & Bentler, 2009). Consequently, not to test statistical model assumptions and not to report their violations are considered "questionable research-related behaviors" in IS research (i. e., two out of 29 proposed by Allen et al., 2011).

During the study of his seminal article, Likert (1932) had observed from the results that "a great number of the five-point statements (...) (in each case the subject being offered five alternatives from which to choose), yielded a distribution resembling a normal distribution" (p. 21). However, the common opinion throughout the literature is that, when analyzing real-world phenomena, chances of actually obtaining normally distributed data are not very high:

- Bentler and Dudgeon (1996) postulated that "in practice, the normality assumption will often be incorrect" (p. 566). In relation to this, they cited a large meta-analysis comparing studies across many different research field which showed that *all* examined studies were affected by significantly nonnormal distributions.

- McDonald and Ho (2002) supported this line of thought when phrasing: "However, (...) much social and behavioral science data may fail to satisfy this assumption" (p. 69). And Gao, Mokhtarian, and Johnston (2008) put it like this: "In general, real-world data (...) do not even have univariate normal distributions, let alone multivariate normal distributions" (p. 116).

- In reference to methods of analysis, B. O. Muthén and Kaplan (1985) therefore postulated that "in practice, factor analysis is often carried out on variables which are highly skewed and/or kurtotic" (p. 171);

- As for the context of MDTs, Enders (2001b) stated that "to date, SEM missing data studies have almost exclusively modeled the unrealistic situation in which the multivariate normality assumption is met".

Many more could be cited here to support the claim that studies with data not suffering from departures from normality are rather rare. In answer to this recurrent observation, authors of high-ranking publications have demanded to relax the assumption of multivariate normality for quite some time, or to even drop multivariate normal theory entirely (e. g., see related work of J. C. Anderson & Gerbing, 1988; and Bernstein & Teng, 1989).

Attempts to address this issue consist of, among others, developing new statistics and methods like the aforementioned ADF approach or bootstrapping techniques, for example, but also so-called *rescaled* (Yuan & Bentler, 2000) or *scaled* (Satorra & Bentler, 1994; Chou & Bentler, 1995; Hu et al., 1992; Miles, 2003) normal-theory test statistics, as implemented in the EQS program[25] (West et al., 1995). Other approaches for data with missing values, for example provided by the Mplus program, are based on Yuan and Bentler (2000)'s work (cf. also Allison, 2003),[26] as well as on Bayesian neural networks like the newly developed universal structure modeling (USM), which has specifically been designed to complement SEM for exploratory research (Buckler & Hennig-Thurau, 2008).

As usual, all of these (more or less new) approaches have their own strengths and weaknesses (cf. Yuan & Bentler, 1999; Fan et al., 1999; Chou et al., 1991; B. O. Muthén & Kaplan, 1985, 1992; B. O. Muthén, Asparouhov, Hunter, & Leuchter, 2011; Finch et al., 1997; West et al., 1995; Ory & Mokhtarian, 2010). Bootstrapping, for example, is liable to distortion if a given sample size is small (Byrne, 2010), a condition held responsible for large standard errors and covariances often not being positive definite (Kline, 2011), and the performance of bootstrapping strongly depends on the parent sample and how it relates to the population sample (West et al., 1995; for an introduction to the technique and associated technical terms, see Hancock & Liu, 2012). ADF is sensitive to small sample sizes, too (West et al., 1995; Bühner, 2011)–especially if distributions are strongly kurtotic (cf. e. g., Olsson, Foss, & Troye, 2003; yet may benefit from a special bootstrap correction, cf. Yung & Bentler, 1994), and is additionally sensitive to degrees of freedom (Yuan & Bentler, 1998, i. e., rather than to model complexity). The widely recommended scaled χ^2 statistic (see above) cannot be used for χ^2 difference testing of nested models, because a χ^2 difference is not distributed as χ^2 (Satorra & Bentler, August 3 / 1999; L. K. Muthén & Muthén, 1998-2010), [27] and especially, it cannot be applied to data with missing values, making it unsuitable for our data requirements. More research will be needed here to extend these approaches. Also, it has been emphasized that measures of goodness

25 Bentler, P.M. (2006). EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc., http://www.mvsoft.com/

26 Muthén, L.K. and Muthén, B.O. (1998-2010). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén. Mplus apparently offers estimation with non-normal data by providing standard error estimates and test statistics that are robust to departures from normality ($ESTIMATOR = MLR$).

27 See also http://www.statmodel.com/chidiff.shtml.

of fit for models estimated by differing methods are incomparable (cf. Fan et al., 1999; Bühner, 2011).

Another example for an estimation method often quoted in the context of relaxed distributional requirements is PLS (Hair, Ringle, & Sarstedt, 2011; Gefen et al., 2011), yet comparative studies have shown the often drawn conclusion that PLS provides notable advantages over CBSEM in the presence of nonnormality to be unsustainable, because (fortunately) both methods "were relatively robust (and equally so) to moderate departures from normality and all suffered (again about equally) to a certain extent under extreme departures from normality" (Goodhue et al., 2012, p. 983). Literature furthermore suggests that generally, "ML estimation is relatively robust in the face of moderate non-normality (...) when large sample sizes are present" (Ory & Mokhtarian, 2010, p. 432; cf. also Chou et al., 1991), and that "ML estimates have been found to be quite robust to the violation of normality. That is, the estimates are good estimates, even when the data are not normally distributed" (Chou & Bentler, 1995, p. 38).

Still, though agreeing that parameters are generally unbiased, some nonetheless advise a certain amount of caution in the presence of nonnormal data, for instance with regard to standard errors (Enders, 2001b; Raykov, 2012), and recommend their correction (cf. Graham, 2009, in the context of MDTs) because rejection rates for model fit induced by nonnormality bias, for example, tend to be "excessive" (Enders, 2001b, p. 354). It has been hypothesized that results from studies on estimation methods assuming normality (like ML-based methods) are generally also valid for MDTs assuming normality (e. g., Enders, 2001a, 2001b), and studies seem to actually confirm this hypothesis (cf. Graham, 2009). Some have therefore offered nonparametric approaches for the treatment of missing data (Paddock, 2002).

In our case, all except one variable (i. e., the player level) had univariate distributions resembling the normal distribution as within the limits found in literature (cf. Section 4.2.3), and the only variable exceeding these limits had no missings at all (see Appendix B, Section B.1 for skewness and kurtosis values). To account for deviations of our performance measure, we performed a transformation, which is a common practice. Also, our sample size can generally be considered sufficiently large to be comparatively such that the ML estimation, for example, should be sufficiently robust against the violation of normality (cf. Ory & Mokhtarian, 2010; Mackelprang, 1970). We discuss the subject of sample size in more detail in a separate section.

Scale Levels, Response Formats, and Scales

During our study, the question arouse whether a particular scale level is inherent to rating scales and/or single rating items, and whether they could (or should) be treated as interval-level or ordinal-level data. In the process of determining these important properties and the adequacy of our measures for different types of analysis, we found the use of terms often confusing and claims with regard to properties of data types contradicting. This was especially true of subjects related to Likert and his work; in 1932, he presented a new technique to measure attitudes, that is, one that was much simpler to implement than those available until then, and which is now one of the most common techniques in empirical research (Weiber & Mühlhaus, 2010). Results of our extensive investigation of literature on the topics of scale level, response formats, and scales are assembled in the list below. We do not make an explicit distinction between agreement ratings and intensity ratings here (see details on our measurements in Section 3.2), as we believe that our findings should apply to both types equally.

- Likert-type rating *items* data is usually considered "discrete and ordinal rather than continuous and of at least interval level of measurement" (Bovaird & Koziol, 2012, p. 497). Bühner (2011) takes the line that, "by definition" (in German: "per Definition", p. 233), the distribution of such items cannot resemble a normal distribution because of their discrete answer format, and that it is not useful to apply a precise normality test to this kind of item data.

- When Likert-type rating *scales* refer to "totals or averages of answers to multiple Likert items" (J. D. Brown, 2011, p. 13), then, as shown in many studies, "Likert scales (as opposed to single Likert response format items) [are assumed to] produce interval data" (Carifio & Perla, 2007, p. 106).

- Representing a continuum is an important conceptual property of rating data. This concept was already introduced by Likert (1932) when he noted: "So far as the measurement of the attitude is concerned, it is quite immaterial what the extremes of the attitude continuum are called; the important fact is that persons do differ quantitatively in their attitudes" (p. 48). An interesting clarification has been given by Carifio and Perla (2007), who argue that when "agree" and "disagree" are the "binary categories of the 'scale' (i.e. response format)", though representing "a severely truncated ordinal response format (and data type)", an underlying continuum can be assumed, as opposed to a situation where categories like "yes" and "no" are used, which represent "a nominal response format (and data type)" (pp. 107–108).

- Estimation problems occur with "four or fewer response options in a Likert-type scale" (Bovaird & Koziol, 2012, p. 497, obviously in reference to a rating *item*), but "when ordinal data based on at least five response options and approximate a normal distribution, normal theory ML chi-square Type I error rates are relatively unaffected" (q. v.); this line of thought also has implications for MDTs (cf. Graham, 2009). As J. D. Brown (2011) remarked: "Allen and Seaman (. . .) support treating Likert scales as interval data with certain rather sensible provisos: . . . . The scale item should be at least five and preferably seven categories" (p. 11).

- In a study, Dawes (2008) examined differences in terms of mean, variance, kurtosis, and skewness that would occur if data on the same construct were gathered using numerical scales either with five-point, seven-point, or 10-point response formats; the data from the five- and seven-point formats were later rescaled to a 10-point format for comparison. When examined, the means produced by the 10-point scale were slightly higher, other differences were not significant.

- Some propose that instruments like rating scales "measure" on a "metrical" level (in German: "metrisch messen" Bortz, 2005, p. 26) or, respectively, "interval scale level" (in German: "Messung auf Intervalskalenniveau" Weiber & Mühlhaus, 2010, p. 99), a proposition based on the "axiomatic measurement theory" (in German: "mit Hilfe der axiomatischen Messtheorie", Weiber & Mühlhaus, 2010, p. 98). From this one can infer that methods of analysis available for metric data which are meaningful to them can be applied (on the subject of "meaningful" in this regard cf. Bortz, 2005; J. D. Brown, 2011; M. C. Edwards, Wirth, Houts, & Xi, 2012; and Bühner, 2011).

- As B. O. Muthén and Kaplan (1985) have observed, "in practice, factor analysis is often carried out on variables which (. . .) frequently are not observed on a continuous, interval scale" (p. 171). Though some suggest that "traditional CFA and SEM assume constructs are measured at least at the interval level of measurement" (Bovaird & Koziol, 2012, p. 495), there is disagreement on whether parametric techniques like CBSEM and PLS actually require an interval scale level at all (cf. Bortz & Döring, 1995, p. 168; and Weiber & Mühlhaus, 2010, p. 99). Studies indicate that using parametric techniques with data which do not perfectly comply with interval scale requirements still provides adequate results (cf. Bortz & Döring, 1995, p. 168).

- Hinkin (1998) states that "Likert-type scales are the most frequently used in questionnaire research[,] (. . .) the most useful in behavioral research [and] (. . .) the most suitable for use in factor analysis" (p. 110).

- In their work on factors and multidimensionality, Bernstein and Teng (1989) point out that criteria "applicable to continuous (*scale-level*) data are therefore inappropriate for discrete (*item-level*) data" (p. 467, parentheses and emphasis in original).

- J. D. Brown (2011) concludes: "Likert scales (...) can be taken to be interval scales so descriptive statistics can be applied, as well as correlational analyses, factor analyses, analysis of variance procedures, etc." (p. 13). It is our experience that software like SPSS will not run if data are imported as anything else than "scale" data for these procedures.

For easily comprehensible reviews and clarification on related misunderstandings, the reader is also referred to Carifio and Perla (2007) and J. D. Brown (2011).

We believe that our measurement approach and the methods of analysis applied to our data are in line with the observations and recommendations listed above; to give an example, our rating scales used a seven-point response format, and the resulting data was only very mildly skewed or kurtotic and approximated a normal distribution.

For completeness, two long known issues need to be mentioned with regard to the use of rating items in particular and in the context of surveys in general, namely the risk of bias caused by differing cultural backgrounds among participants (cf. Likert, 1932) and by issues related to social desirability (cf. D. L. Phillips & Clancy, 1972).

Future approaches may break new ground to present information in questionnaires and to collect data from survey participants, like additional visual boundary lines and evaluative labels to a graphical format (Peters et al., 2009), for example, as research promotes our understanding of how humans process information during surveys especially with regard to "feelings" and "thoughts" (q. v.) and may deliver valuable new insight into our research topic.

General Issues

This section discusses issues we came across when dealing with our general research approach which seldom receive the attention they deserve when studies are being published. At the end, we present alternative approaches which could also lead to interesting findings in terms of the purpose of the study.

Sample Size and Power Analysis

A small sample size is problematic even for normal and complete data (Yuan & Bentler, 2000). As to the issue of sample size with regard to model testing, there seems to be no generally accepted definition of a "small" or "large" sample (cf. Marcoulides & Saunders, 2006, p. iii; Kline, 2011, p. 11). Some clues are yet provided in publications:

- In the context of misspecification of factor-indicator relationship types in interactions with sample size, Roberts et al. (2010) have differentiated between sample sizes of 250 as a small and of 500 as large.

- Kline (2011) considered sample sizes of 200-300 “typical in SEM” (p. 201).
- West et al. (1995) stated that large samples are “clearly in the range of 1000 to 5000 cases” (p. 74) in the context of SEM.
- Finch et al. (1997), Fan et al. (1999), and Hox, Maas, Cora J. M., and Brinkhuis, Matthieu J. S. (2010) have tested the impact of sample size on statistics with *simulated* samples of sizes starting from 50 up to 1,000; similarly, Bearden et al. (1982) experimented with sizes from 25 to 10,000.
- After evaluating and comparing PLS practices published in MIS Quarterly on the one hand and in Journal of Marketing, Journal of the Academy of Marketing Science and Journal of Marketing Research on the other, Ringle et al. (2012) reported sample size means of 238.12 (ranging from 17 to 1,449, median: 198, $n = 109$) and 210.88 (ranging from 39 to 2,990, median: 160, $n = 160$), respectively.

The aforementioned numbers from different sources may indicate that, in comparison with other studies, our sample size ranged among the larger ones.

As elaborated above, sample size is of great significance when dealing with nonnormal data, but the issue of sample adequacy particularly raises when statistical power of an analysis needs to be planned (cf. Lai & Kelley, 2011) or investigated. The minimum sample size of a study is not invariant across studies, and whether a sample size is adequate or inadequate can thereby not be determined independently from several aspects of a particular study (MacCallum, Widaman, Zhang, & Hong, 1999). Often cited and applied rules of thumb to determine the required sample size have been found unsuitable guidelines[28] as they typically lack empirical substantiation (Lai & Kelley, 2011; Goodhue et al., 2012; In’nami & Koizumi, 2013). On the one hand side, a given sample size can be inadequate by being too small, but on the other hand side also by being too large: If, for instance, for a predetermined effect size, a sample size is too large, even unsubstantial effects will become significant (Bortz & Döring, 1995; Kline, 2011). However, as small sample sizes are much more common, issues like increased standard deviations, decreased statistical power, and reduced accuracy related to an insufficient sample size (e. g., Finch et al., 1997; Goodhue et al., 2012; K. H. Kim, 2005; for examples, see Bernstein & Teng, 1989; Ringle et al., 2012) have been studied much more frequently then issues related to the inverse situation, and only few works have examined problems the consequences of small and large samples sizes simultaneously (e. g., Ory & Mokhtarian, 2010).

28 See Goodhue et al. (2012) for examples of recent articles in top-ranking journals which cite such rules of thumb and a criticism of this practice, and Bagozzi and Yi (2012) for a contrasting point of view on the subject.

While some studies have shown the relative robustness of the PLS algorithm against small sample sizes (Majchrzak, Beath, Lim, & Chin, 2005), thereby giving the impression that it is sufficient to rely on the algorithm to deal with the sample size issue, others have concluded that PLS and ML are likewise subject to substantial bias such as "increased standard deviations, decreased statistical power, and reduced accuracy" when applied to small sample sizes (Goodhue et al., 2012, p. 981). Bühner (2011) emphasized that the χ^2 test, which is reported "in virtually every application" of CBSEM (MacCallum et al., 1996, p. 132), actually tests the null hypothesis and that one is conceptually dealing with a Type II or β error in this context, but that predefining an effect size–as required for a correct interpretation test statistics–is not a common practice in SEM studies (cf. also Barrett, 2007). In the same line, Cohen (1994) quoted one of his colleagues who

> reminded researchers that, given the fact that the nil hypothesis is always false, the rate of Type I errors is 0%, not 5%, and that only Type II errors can be made, which run typically at about 50% .... He showed that typically, the sample effect size necessary for significance is notably larger than the actual population effect size and that the average of the statistically significant effect sizes is much larger than the actual effect size (p. 1000).[29]

Table 20 reproduces the four possible outcomes of a study with regard to rejecting or not rejecting the null hypothesis H_0. A very good discussion of the correct interpretation of statistical significance is also given by Kline (2011, pp. 36–39).

	True State	
Decision	H_0 True	H_0 False
Do not reject H_0	Correct decision $(1-\alpha)$	Type II error (β)
Reject H_0	Type I error (α)	Correct decision $(1-\beta = \text{Power})$

Table 20: The four possible outcomes in a study and their probabilities. Adapted from Kim, K. H. (2005). The Relation Among Fit Indexes, Power, and Sample Size in Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal, 12(3), p. 369.

29 At this point it is to be noted that the estimation of effect size parameter, the standardized difference between two population means, also assumes normality, as well as homoscedasticity (i. e., homogeneity of variances in the therapy and the control group, see Grissom & Kim, 2001). See also Lai and Kelley (2011) on the importance of confidence intervals for the model parameters of interest related to SEM and sample size planning.

Nonetheless, as W. W. Chin (1998) notes, "there continues to be a neglect of statistical power analysis in the behavioral sciences" (p. xi), a particularly incomprehensible behavior if exhibited by editors and reviewers (Cohen, 1992). K. H. Kim (2005) stresses that addressing the power of a study and the requisite sample size to achieve that power is important for any kind of research, SEM included, and Hoyle and Isherwood (2013) presents a long list of sources of recommendations for various disciplines. Moreover, a variety of statistically sound techniques to determine appropriate sample size and/or statistical power are available which are explicitly recommended.[30] For example, Goodhue et al. (2012), Bortz and Döring (1995) and others have recommended the approach provided by Cohen (1988), regardless of the subsequent analysis procedure envisaged. Other suggestions for the estimation of the requisite sample size for a particular model and the evaluation on the adequacy of the size of a sample include the generation of complementary measures (Bearden et al., 1982) of conduction of additional analyses (e. g., Monte Carlo procedures, see In'nami & Koizumi, 2013). Cohen (1994) recommended to report "effect sizes in the form of confidence limits. 'Everyone knows' that confidence intervals contain all the information to be found in significance tests and much more" (p. 1002, cf. also Lai & Kelley, 2011). It is possible that given a particular null hypothesis, different alternative hypotheses may result in the same level of power for a test, which implies that the same effect size can be obtained for differing alternative hypotheses (MacCallum, Lee, & Browne, 2010; cf. also Oertzen, 2010). In the SEM context, T. Lee, Cai, and MacCallum (2012) have suggested to determine the necessary sample size using a RMSEA-based approach (MacCallum et al., 1996, note that the respective fit values are derived from literature but some of them are somewhat arbitrary. p. 135), a power analysis and hypothesis tests for fit of single covariance structure models which has later been extended for evaluating nested models (MacCallum, Browne, & Cai, 2006); for a discussion on issues related to hierarchical or nested models and equivalence, see Wetzels et al. (2009) and Bentler and Satorra (2010). As the distribution of the RMSEA index is known, its "degree of imprecision" (MacCallum et al., 1996, p. 130) can be accounted for through a confidence interval (cf. F. Chen et al., 2008). With this approach, the effect size not expressed as a numerical index, but is instead determined in terms of the difference between a null value ϵ_0 and alternative value ϵ_a of the RMSEA.

Because this approach has specifically been designed for our particular requirements and because of its excellent characteristics, yet also by reasons of simplicity and ease of use especially compared to other

30 Interestingly, this recommendation was not always followed by those who gave it in the first place (cf. Goodhue et al., 2012); in the study by Majchrzak et al. (2005) of which W. W. Chin (1998) was a coauthor, the sample size was justified by the so-called "5 to 10 times" rule (p. 660) rule and the power analysis consisted of an ad hoc simulation, very briefly described and of unverifiable explanatory value.

methods (D. Kaplan, 1995), we decided to apply the latter in order to determine the statistical power of our CFAs. K. H. Kim (2005) provide a step-by-step demonstration of how to calculate the required sample size on the basis of this (though values in and cited from the original paper appear to have been mixed up or to differ). A special software, accessible per web interface,[31] facilitates the application of the MacCallum et al. (1996) approach. It automatically generates corresponding program code for the R programming environment[32] using the parameters predefined by the user. Once pasted into an R console and run, the code allows for the generation of estimates of the statistical power. An assessment of the minimum sample size required to achieve a given level of power, of the power for testing the difference between two nested models (cf. MacCallum et al., 2006), or of the minimum sample size required to achieve a given level of power for a test of nested models can also be provided. Before actually using the software for our study, we tested it with values from example tables of the original paper and compared its outcome with the published ones to ensure the software was working correctly.

It is noteworthy that while the determination of ϵ_0 and ϵ_a values roughly determines the level of fit for the single model test, the choice of suitable RMSEA values for comparing models is assumed to have little impact on power regardless of any other influence factor if the sample size is very large (MacCallum et al., 2006)–like ours was.

Model Specification and Identification

When intending to use SEM it is recommended that model specification should best be accomplished before the selection of the final measures (and therefore before data collection, screening, and preparation), a recommendation which we complied with in this study. When using PLS (i. e., variance-based SEM), so-called model identification is not an issue (Weiber & Mühlhaus, 2010; Roberts et al., 2010), but if using CBSEM–for instance when developing of new measures–also model identification should precede data acquisition (Kline, 2011; MacKenzie et al., 2011). If checked for identifiability later, the final CBSEM model will be restricted by the variables available in the data and the circumstances of data collection (Hoyle, 2012c)–however, this scenario is not uncommon across research disciplines (q. v.).

The problem with model specification is that determining whether a model is identified can be a very difficult task (MacCallum, 1995). Models meeting the minimum condition of identifiability may still not be identified, so that it has even been suggested the problem of model identification is *undecidable* (Kenny & Milan, 2012, p. 149). However, standard CFA models will be identified if they meet certain requirements

31 Preacher, K. J., & Coffman, D. L. (2006, May). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://quantpsy.org/.
32 The R Foundation for Statistical Computing, www.r-project.org/.

(Kline, 2011). For reasons of time constraints between the pretests and the final data collection, a possibly required test of identifiability of our complete hypothesized SEM was scheduled to such that it would be performed after the data collection.

Factor-Indicator Relationships and Indicators per Factor

The necessity to distinguish between different types of indicators for latent variables, that is, indicators either being a cause or an effect of a construct (Blalock, 1963), has first raised researchers' awareness some decades ago; the two types of indicators, causal and effect indicators, are often called *formative* and *reflective*, respectively (Jarvis, MacKenzie, & Podsakoff, 2012), although these terms are used ambiguously (Bollen, 2011).[33] The different types of models which can be distinguished on the basis of the directions of their factor-indicator relationships–relationships between manifest and latent variables, but also between latent variables and a higher-order variable–have been given various different names by different authors (see Wetzels et al., 2009, for an enumeration; also MacKenzie et al., 2011).

The importance of modeling factor-indicator relationships adequately, or, in other words, determining the type of measurement model correctly, has been stressed in the literature of different disciplines, such as psychology, marketing, and IS (cf. e. g., MacKenzie, Podsakoff, & Jarvis, 2005; Diamantopoulos & Winklhofer, 2001; Bollen, 2011). The discussion about the consequences of incorrect measurement model specification is ongoing to date (e. g., Bagozzi, 2011; Jarvis et al., 2012; Cenfetelli & Bassellier, 2009): While it is generally agreed that the impact of the relation between a construct and its indicators on theorizing and testing is profound (Wong et al., 2008) and that specifying this relation incorrectly may possibly affect parameter estimates (MacKenzie et al., 2005), also in the context of MDTs (Black et al., 2011), the actual effect of misspecification with regard to different parameters may be less severe than previously assumed (cf. e. g., Aguirre-Urreta & Marakas, 2012; Jarvis et al., 2012; Petter, Rai, & Straub, 2012). Whether a construct is measured in a formative or reflective fashion can also amplify the impact of missing values on reliability and validity of results. For instance, if data points for indicators are missing, the impact should be larger in the case of formative measurement because all components of a formative construct are critical with regard to its conceptualization (McKnight et al., 2007). In contrast, for reflective constructs, the indicators are considered interchangeable and are expected to point into the same direction (Jarvis et al., 2003), so that if data for a reflective indicator is missing, the "gap" may be filled using the ones remaining for the construct.

33 For example, "the term *formative indicator* came much later than causal indicators and had a meaning more restrictive than causal indicators" (Bollen, 2011, p. 360, emphasis in original)

Constructs are yet not *inherently* formative or reflective, and whether they are adequately modeled as having formative or reflective indicators should be inferred from their conceptualization (MacKenzie et al., 2011). Also, a measurement model is not required to be purely formative or reflective; for example, higher-order constructs may have first-order subdimensions as formative indicators which themselves may have reflective indicators (cf. MacKenzie et al., 2011). To help determining the type of measurement model of a construct correctly, researchers have generated and repeatedly pointed to guidelines covering coherent sets of conceptual criteria which lead a researcher through the whole decision process (e. g., Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003; Bagozzi, 2011). In line with recommendations obtained from the abovementioned literature, we have developed a procedure which–to the best of our knowledge–was suitable to identify the appropriate type of measurement model for our construct conceptualizations. This procedure has been demonstrated in Section 3.2.1, where we also explain our reasoning behind every step.

The type of measurement model also determines the amount of indicators that should be used to measure a construct. In the case of formative measurement, due to the concept of a construct which is formed by its *composite* indicators (cf. Bollen, 2011), only multiple-item measurement seems appropriate (W. W. Chin, 1995), while for reflective measurement, both single- and multiple-item measures are considered suitable[34] (Weiber & Mühlhaus, 2010). Studies in the context of CFA performed on the basis of ML (i. e., in the context of reflective measurement, cf. Diamantopoulos & Winklhofer, 2001) have demonstrated the importance of the number of variables defining a factor for the purpose of obtaining stable parameter estimates (Bernstein & Teng, 1989). More indicators per latent variable produce more exact and more suitable results, more reliable factors, and less often lead to estimations that do not converge (Bühner, 2011). Likert (1932) himself pointed out that "a sufficient number of statements should be used in each form to obtain the desired reliability" (p. 51). The recommended number of needed indicators per construct typically ranges from three to six (cf. Hinkin, 1998; Weiber & Mühlhaus, 2010). In a meta-analysis conducted by Ringle et al. (2012) observed means for numbers of indicators found in actual studies were 3.58 for reflective and 3.03 for formative constructs. With reference to the required number of indicators, all our measures followed the recommendations above except for performance (being a construct represented by a single indicator, cf. Kline, 2011, p. 119); but, as we could assume that our performance indicator would measure without error, that is, with limited ambiguity (cf. Gefen et al., 2011; Weiber & Mühlhaus, 2010), we considered this approach to be justified.

34 Though some strongly discourage single-item measures per se and even call it a "folly", like (Churchill, 1979, p. 66).

However, details of the indispensable and informative decision processes related to measurements are seldom given in publications of actual studies. If at all, authors confine themselves to short remarks, mentioning the type of measurement in passing. Burton-Jones and Straub (2006), for instance, only briefly comment that "to measure cognitive absorption, we adopted Agarwal and Karahanna's (2000) prevalidated scale .... [Agarwal and Karahanna] model cognitive absorption as a reflective, higher-order construct" (p. 237), whereas the original work by Agarwal and Karahanna (2000) is even shorter in length: "All of the constructs were modeled as reflective and most of the constructs in the model were measured using multiple indicators" (p. 683). This brevity is remarkable because to this day, "a significant number of articles have misspecified (...) constructs" (Bagozzi, 2011, p. 269), and information provided by some fit indexes appear to be spurious for misspecified models (Fan et al., 1999). One would assume that high-ranking journals would insist that submitting authors make their decision process on the type of measurement model available to their readership.

Factor Extraction Methods, Rotation, and Extraction Criteria

It is well established that the methods used for extraction and rotation of factors during EFA can seriously affect the quality of analysis results (Grice, 2001). In contrast to other techniques which have been heavily criticized, we chose to use ML and PAF which are widely recommended (Patil et al., 2008; Weiber & Mühlhaus, 2010).

Various authors have furthermore emphasized the importance of the decision on how many factors to retain during an EFA, and some argue that this decision has more impact on the results of the analysis than any other selection of method to be made during the EFA process (Rudner, 2007). Many have found fault with the practice of using common rules of thumb like, for example, the so-called "eigenvalue greater than 1.0" rule, for determining the number of factors to be retained, especially if researchers intend to rely on one retention criterion only (cf. Conway & Huffcutt, 2003). Conversely, Horn's PA and Velicer's MAP test are highly recommended (e. g., Bühner, 2011; Rudner, 2007; Timmerman & Lorenzo-Seva, 2011), because both are statistically sound and therefore superior to other techniques. They "typically yield optimal solutions to the number of components problem" (O'Connor, 2000, p. 396; for a comparison with, e. g., Cattell's scree test and Kaiser's "eigenvalue greater than 1.0" rule, see Zwick & Velicer, 1986; Patil et al., 2008). Both techniques were originally grounded on PCA: MAP, for example, involves a PCA of the correlation matrix of variables (Patil et al., 2008). Some effort has been made to extend them in such a way that they can now be used on the basis of a PAF computation as well.[35]

35 The debate is ongoing on whether to base an extraction on communalities or principal component eigenvalues, respectively, and the issue remains unsettled; see supplemental material of O'Connor (2000): SPSS, SAS, and MATLAB Programs for

Depending on the data in hand, one combination of techniques (PA vs. MAP, either in the form of a PCA or a PAF) may outperform the respective other (Rudner, 2007; Crawford et al., 2010; Bühner, 2011; Timmerman & Lorenzo-Seva, 2011). The original MAP test of 1976 has been revised in 2000; we indicated results of both tesct for the primary constructs in the respective table Table 14.

Scale Refinement

Because they can usually be assumed to be relatively reliable (see Bühner, 2011, p. 314), we favored to use existing scales where possible (cf. Bagozzi, 2011). To be on the safe side and to avoid any translation or language issues, we yet assessed and revised every candidate item regardless of their source, in order to refine and purify all the measures (MacKenzie et al., 2011). Of course, we draw most attention to the newly developed CIRME construct: We (quite rightly) assumed that a large amount of items in the pilot study questionnaire could deter our survey candidates of even before they had had the chance to answer the CIRME items, which would have reduced our chances of gaining as much experience with these items as possible *before* the final survey. A pilot study with a smaller questionnaire only containing CIRME items was a way to maximize our yield in this regard: In contrast to Pilot 2 which covered all constructs and controls (i. e., all other candidate items that had been developed to far), Pilot 1 only covered our self-developed CIRME items.

As a matter of fact, the samples from our pilot studies were much smaller than the one obtained from the final data collection. Also, the variable the distributions of the pretest data were not well suitable for the application of ML methods and additionally, we encountered typical estimation problems like a nonpositive definite matrix (cf. Wothke, 1993; Kline, 2011), possibly because our pretest sample was too small, because no inverse matrix could be build, or because the determinant of the matrix became zero during due to missing values, outliers, or nonnormal data (Bühner, 2011). These circumstances limited the options we had in terms of MDTs and methods of analysis considerably, and we had to make compromises.

So as far as the treatment of missing values is concerned, a problem with (recommended) MDTs which are based on EM or ML is that their algorithms frequently fail to converge due a nonpositive definite correlation matrix (Graham, 2009), as happened with in our pretest data. We therefore had to fall back on the less favorable ad hoc procedures, out of which we chose to apply listwise deletion.

As for analysis, instead of using more sophisticated techniques like PAF or ML factor extraction, we applied a PCA for the purpose of reducing the number of our observed variables (Conway & Huffcutt, 2003),

---

Determining the Number of Components and Factors–Using Parallel Analysis and Velicer's MAP Test, https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html.

that is, to represent them in as small a number of dimensions as possible (Floyd & Widaman, 1995) or to "suggest dimensions" (Churchill, 1979, p. 69), again due to characteristics of our pretest data. The discussion about PCA vs. PAF is ongoing: Both are typically seen as exploratory procedures of factor analysis (Floyd & Widaman, 1995), yet some consider PCA not a factor analytic method as such (Bühner, 2011), others do (Bortz, 2005), and again others prefer PAF over PCA in general because the latter mixes "common, specific, and random error variances" (Hinkin, 1998, p. 112). However, Conway and Huffcutt (2003) observed that in the context of psychology and organizational research PCA "predominated over common factor analysis" and that the results even indicated "a trend toward increasing use of PCA over common factor analysis" (p. 151) in the period investigated (i. e., from 1975 to 1999). In approximately a quarter of the examined publications, this issue was not even paid attention to, as the "factor extraction model was not stated" (q. v.). We were also able to find examples of the practice of applying PCA in high-ranking journals: In their study on a proposed consensus on appropriation (COA) scale, Allport and Kerler (2003) used "two separate sets of data" (p. 356), one which contained the initial items of the COA scale which were analyzed via PCA, and another one which contained a revised set of COA items which were analyzed via CFA and later SEM. Again, the use of PCA was not particularly mentioned or justified. Also, although listwise deletion does have its deficiencies, we believe that this was an acceptable MDT at the early stage of refinement and purification and for this particular purpose. Sufficient evidence for our belief can be found in literature: For instance, though advising against pairwise deletion in general, Graham (2009) considered this method useful "when conducting preliminary exploratory factor analysis with a large number of variables" (p. 554) and when the focus does not lie on the results of the factor analysis itself. Roth (1994) even explicitly recommended to either use listwise or pairwise deletion to factor analyze data, in line with Mackelprang (1970). The treatment of missing values in the final data was discussed above.

At this point it is worth noting that scale levels of measures (see item 5.3) also has implications for results of a factor analysis; Bernstein and Teng (1989) suggested that "any present method to determine dimensionality of an item set will exaggerate the number of factors needed if one applies the same criteria to these data as they would to continuous data" (p. 476). During all our factor exploration activities, we thus kept in mind that if in doubt, as "parsimony and simple structure are desired for the scales" (Hinkin, 1998, p. 112), we should possibly consider a less complex model. Dropping a subdimension yet always contains the potential risk of "eliminating an essential aspect of the construct domain" (MacKenzie et al., 2011, p. 316), as we experienced with self-motivational traits.

Validation, Estimation, and Model Fit

Differing opinions concerning the suitability of measure purification and validation techniques have been expressed. MacKenzie et al. (2011), for example, consider Cronbach's α as an important information in this regard: "Cronbach's alpha has traditionally been used to estimate the internal consistency reliability of the measures. This is appropriate" (p. 314). However, as Cortina (1993) remarked, "although alpha is sometimes referred to as 'the' estimate of reliability, it is not the only estimate of reliability" (p. 98). To compare the model fit of a hypothesized factor structure of a construct to the fit of a concurrent model for that construct, some instead propose to apply a CFA and to use related measure to examine reliability (Byrne, 2010). Despite the fact that such analyses are still requested by many journals, some even consider the analysis of reliability based on Cronbach's α and similar measures unnecessary or redundant in the SEM context. They argue that the information supplied factor loadings and error variances "incorporates reliability" (Bagozzi & Yi, 2012, p. 16), making a CFA the more "rigorous" approach (same source, p. 14). Again others have contrasted this view: Patil et al. (2008), for example, have cautioned that "skipping" EFA in favor of an "exploratory" use of CFA and the practice of modifying hypothesized items and/or constructs bears certain risks of misjudging the underlying number of factors in data. However, when applying EFA, it is important to follow certain guidelines (Conway & Huffcutt, 2003). To meet all requirements and the highest standards, we applied several techniques to investigate the psychometric properties of our measures and used both types of factor analysis as recommended (e. g., Bernstein & Teng, 1989; MacCallum et al., 1999).

After having evaluated the various fit measures for each construct, the validity of indicator sets at the construct level as well as the reliability of indicator sets at the construct level (cf. MacKenzie et al., 2011; Bühner, 2011), or, to summarize, after testing the validity of the measurement model of all constructs integrated in our hypothesized model (cf. Byrne, 2010), we were prepared for testing the causal relationships we had postulated in by means of SEM. Yet in order to do so, we first had to overcome several obstacles. Due to the distribution of our dependent variable and to the fact that our data did not follow a multivariate normal distribution, we considered using estimation methods other than ML. Though some do not seem to consider this a true SEM approach (cf. Hair et al., 2011; Kline, 2011; cf. also MacKenzie et al., 2011; Bagozzi & Yi, 2012; Hammervold & Olsson, 2012, who do not even mention PLS), many have proposed to use PLS in theory building stages in the past because it is "explicitly designed to establish that relationships exist and explain meaningful amounts of variance" (Roberts et al., 2010, p. 4331). Another reason for preferring this approach is the fact that PLS makes it easy to deal with formative scales, which "presents challenges in CBSEM" (Gefen et al., 2011, p. vi). However, it is generally acknowledged that PLS is "consistently less accurate" (Goodhue et al., 2012, p. 981) than CBSEM, that it does not compensate for measurement error, thereby yielding biased parameter estimates (Gefen et al., 2011), and that it lacks global indexes of fit (Wetzels et al., 2009; Weiber & Mühlhaus, 2010). Also, it is not as robust against small sample sizes and nonnormal distributions as generally assumed; in fact, it not superior to CBSEM in this regard (W. W. Chin, 1998; Marcoulides & Saunders, 2006; Marcoulides, Chin, & Saunders, 2009; Goodhue et al., 2012). A issue seldom addressed in this context, yet important with regard to our study, is that of the adequate treatment of missings when using PLS. As a matter of fact, when using SmartPLS, the user is offered a choice between listwise deletion and mean substitution, two approaches with weaknesses we have discussed in length above. It has been seriously doubted that the PLS algorithm is able to cope with missing values such that proper results are to be expected (Parwoll & Wagner, 2012), hence it was inappropriate in our case.

With regard to model fit, it has been suggested to interpret related measures in the context of "the dichotomous decision process of hypothesis testing: The model was either accepted as providing good fit to the data, or the model was rejected as fitting the empirical data poorly" (Fan et al., 1999, p. 57). We therefore followed the recommendation to always rely on several fit measures to adjudge a model (F. Chen et al., 2008). Well-known expert in the SEM field encourage research to embrace the results of their studies irrespective of the fact whether their initial hypotheses are rejected or confirmed, stating that obtaining unexpected results is much more interesting and that there is no "shame" in retaining the originally hypothesized model (Kline, 2011, p. 203). On the other hand, authors equally demand to acknowledge the fact that alternative and equivalent or near-equivalent models may exist which have the same power of test of fit (MacCallum et al., 2012; Beier & Ackerman, 2005; K. H. Kim, 2005), and that even if they are found, the question of model selection is not trivial (Preacher & Merkle, 2012; Bühner, 2011). One recommended way to make a decision is to test whether a model is scientifically meaningful, no matter how well the model fits the data (Kline, 2011; Freedman, 1991). Though altering a model according to so-called modification indexes in order to achieve better model fit is common practice (Byrne, 2010), many argue that "this approach is notorious for problems and failures, including invalidity of modifications (...) and capitalization on chance" (...) (MacCallum et al., 2012, p. 344) and dismiss this practice of in general (MacCallum et al., 1992; W. W. Chin & Todd, 1995; W. W. Chin, 1998).

Alternative Methods of Analysis

For the present study, we chose SEM as our general research approach, which we believe to be appropriate for our research question. Also, in order to ensure highest reporting standards, we carefully reviewed related literature and applied some of the most scientifically rigorous techniques available. However, as with any other methods of analysis, SEM is not without its limitations (Barrett, 2007). On a more general note, one should remember "that a construct is simply a concept created for scientific purposes.... As it does not physically exist, there can be no 'true' conceptualization of a construct, and, for most purposes, any construct is as good as any other" (Barki et al., 2007, p. 188). Also, constructs could be represented through alternative methods than the one we chose (Bagozzi & Edwards, 1998). Also, all goodness-of-fit measures depend on specific conditions (Beauducel & Wittmann, 2005; F. Chen et al., 2008). Moreover, comparing the default model to the so-called null (or independent) model can be inappropriate, such that these indexes allow for no meaningful interpretation (Widaman & Thompson, 2003). Additionally, it is also important to know that the conceptualization of the null model may differ across different software or even across different versions of the same software, with the consequence that "a large number of published studies (...) may have come to grossly inaccurate conclusions" (Gignac, Palmer, Bates, & Stough, 2006, p. 144).

Numerous new approaches have been proposed to deal with various issues which have been identified so far: Coarsely categorized variables (West et al., 1995), latent class analysis (Vermunt, van Ginkel, van der Ark, & Sijtsma, 2008), exploratory SEM (Marsh et al., 2009, 2010; Asparouhov & Muthén, 2009; B. O. Muthén et al., 2011; Kline, 2011), assessment of scale validity through video techniques (MacKenzie et al., 2011; N. P. Podsakoff, Podsakoff, MacKenzie, & Klinger, 2013), new cross-validation methods for scales (MacKenzie et al., 2011; W. W. Chin & Todd, 1995) and new techniques to assess content validity (Hinkin, 1998; Hinkin & Tracey, 1999; MacKenzie et al., 2011; Gefen et al., 2011). Our measures could also profit from an additional multitrait-multimethod assessment, or by conducting laboratory experiments (Adams et al., 1992, cf.). As Schultze and Orlikowski (2010, p. 812) phrased with particular reference to the context of our study:

> The distinctive characteristics of virtual worlds, however, pose a number of significant theoretical and methodological challenges for the field. On the theoretical side, it is unclear whether existing theories are able to effectively explain the complex and dynamic interactions and events that unfold in real time within the persistent environments that are virtual worlds. On the methodological side, established techniques of social science research such as interviews, observations and surveys may not effectively capture the novel practices that constitute virtual worlds (p. 812).

During our study, we indeed came across the idea that "many cognitive processes underlying human behavior are not accessible to consciousness, and thus not open to introspection and self-reporting, our investigation" (Riedl et al., 2011), or that the differences between individuals may deeply be rooted in biological mechanisms.

As with regard to (more or less) new developments in psychometric measurement, item response theory and Rasch analysis need to be mentioned (Bernstein & Teng, 1989; McKnight et al., 2007; Wright, 1996). Particularly, approaches combining item response theory (IRT) with SEM seem promising (Ferrando, Anguiano-Carrasco, & Demestre, 2013; Glockner Rist & Hoijtink, 2003), and "a number of authors have shown how Rasch analysis can be used to analyze and improve Likert scales as well as transform them into *true* interval scales" (J. D. Brown, 2011, p. 11, emphasis in original). There are also alternatives to factor analysis (Kiang & Kumar, 2001), further data-driven automated strategies (Marcoulides & Ing, 2012), and truly exploratory approaches based on Bayesian neural network like USM (Buckler & Hennig-Thurau, 2008; Weiber & Mühlhaus, 2010). Various R packages have been developed by now, offering exciting opportunities for new discoveries (Gignac, 2013; Fox, Byrnes, Boker, & Neale, 2012). We expect that exciting findings should emerge from applying the methods of analysis listed above to research questions related to virtual worlds.

This chapter has evaluated and interpreted the results and discussed possible limitations of our study. The upcoming discusses implications of our findings on higher level of abstraction, points to possible future research questions, and presents concluding remarks.

6 CONCLUSION

When you have exhausted all possibilities,
remember this: you haven't.

— *Thomas A. Edison*

The previous chapter has summarized our findings with regard to our hypotheses and discussed possible limitations of our study; furthermore, it has discussed important aspects of analysis which relate to missing values and other data issues. This chapter discusses our contribution and infers possible implications of our findings for various areas of application. It also addresses open issues and proposes opportunities for future research. It concludes with final remarks.

6.1 CONTRIBUTION AND IMPLICATIONS

At this point, we would like to refer to the goals we set ourselves in the introduction. We stated that we wanted to

1. contribute to a better understanding of individual performance in virtual worlds and
2. create awareness for missing data in order to promote standards for their treatment.

The results of our study indicate that different requirements are needed for accomplishing different tasks: while some may involve specific emotional capabilities, some may not (cf. also Joseph & Newman, 2010). From our standpoint, this is the major contribution of this study with regard to our first research goal, as these findings clearly demonstrate the importance of well-defined performance measures and lack thereof. As set out in detail in Chapter 2, the use of the concept of performance in literature is, on the one hand, very broad and unspecific in terms of conceptualization, on the other hand very narrow and specific in terms of measurement. Work-related performance, for instance, has traditionally and naturally been of high interest to organizational and societal research. However, neither conceptualizations nor measurements are comparable across various studies. Definitions of work-related performance range from, for example, supervisor ratings, training success, tactical navy decision making, time to completion for a decision-making task, perceived performance impacts, idea generation, promotion, and realization, verbal intelligence and scores in a science quiz to the number of words accurately transferred from an article to a spreadsheet during an allotted time (J. E. Hunter, 1986; Côté & Miners, 2006; B. S.

Bell & Kozlowski, 2002; S. T. Bell, 2007; Goodhue & Thompson, 1995; Janssen & Yperen, 2004; Brackett et al., 2006; Ehrlinger & Dunning, 2003; Carton & Aiello, 2009); furthermore, performance measures have been linked to such complex mediators as proactive behavior, degree of autonomy, and self-efficacy (T.-Y. Kim et al., 2009; Judge, Jackson, Shaw, Scott, & Rich, 2007). Even though general cognitive ability may account for a significant proportion of the various performance measures just listed: It is difficult not to agree that these means of measurement likewise cover very different qualities *in addition* to cognitive ability, and that therefore, the set of abilities required for high achievement with respect to these means will most likely differ, too. Other issues refer to the question of objectivity. In this regard, supervisor ratings are a very good example: It has been argued that in order to compare performance scores across different types of working tasks and to be able to generalize from findings, supervisor ratings are the measure of choice; they are typically available for all participants of a study and "can be used for virtually any type of job, including jobs in which objective performance is difficult or impossible to measure" (Côté & Miners, 2006, p. 11). Yet it is part of everyone's experience that mutual affection, interpersonal differences, the pursuit of differing ends as well as differences in cognitive styles and working approaches–to name a few–significantly affect the supervisor-subordinate relationship, and it would be surprising if these factors would not influence supervisor ratings. For such a measure to be objective, researchers will have to put a lot of effort into controlling for effects like personality traits as well as numerous other effects.

In our opinion, work-related performance will most certainly play a key role among the relevant outcomes of virtual worlds' application to business contexts. This similarly applies to what has been labeled "MIS success", a concept which has broadly been described as capturing individual impact or organizational impact (Goodhue & Thompson, 1995). Agreement on what work-related performance (or any other performance or success measure) should include or exclude as well as clarification of what aspects of performance research has measured so far yet remains to be attained before tapping any further into this concept; the development process for objective cognitive ability measures may serve as an example of how to achieve this goal.

Concerning the use of self-estimated or self-reported measures of cognitive ability, we suspect that interdependencies between institutional achievement feedback and the abovementioned aspects (mutual affection, etc.) may exist, too. Our results furthermore lead to the conclusion that institutional achievement feedback may reflect social competence, as the link to emotional capabilities was remarkable.

The second important contribution with respect to individual performance in virtual worlds lies in substantiating the perception that drivers of IS use may not be of much predictive value with regard to performance (Lucas, 1975), or alternatively, suggesting that usage measures like cognitive absorption may affect performance, but not in every situation (cf. differing results obtained by Burton-Jones & Straub, 2006).

It has been claimed that "indeed, there is widespread agreement among researchers that system usage is *the* primary variable through which IT affects white collar performance" (Straub & Limayem, 1995, p. 1328, emphasis in original). However, to support this claim, we need to turn our attention to ways how to objectively measure performance such that these measurements can be replicated–this is where we come back to our initial point. Though we would not go as far as to fully agree with Benbasat and Barki (2007), we believe that there is an element of truth in positing that "the intense focus on TAM has diverted researchers' attention away from other important research issues and has created an illusion of progress in knowledge accumulation" (p. 211).

Regarding our second research goal, we think that by sharing the experiences of our particular venue, we made a valuable contribution to creating awareness for issues related to missing values and other data issues. To this end, we showed that obtaining guidance from literature–or, related therewith, reassurance–in terms of our own specific missing data situation was hard and arduous. We further explained that it can be challenging to interpret recommendations, especially if they are equivocal or appear to be–or actually are–contradictory. We presented the difficulties it causes to transfer subtle hints from casual remarks and half sentences to real-life missing data problems. We also demonstrated that actually *applying* recommendations is even more difficult, especially when suitable processes are not available. Moreover, we illustrated that all our choices of methods of analysis and effectively all indexes of goodness-of-fit or tests of significance were affected by the fact that our data contained missing values (and that our sample was very large, too). As a result, many workarounds were needed, making the development of a coherent approach a troublesome task.

In this light, it is not surprising that researchers may still use inappropriate MDT–a practice essentially criticized by all authors dealing with missing values–or may prefer not to mention their treatment of missings in a publication. During the course of our study, it often occurred that recommendations lay stress on "Don'ts" rather than providing "Do's"; also, if recommendations treat one specific aspect of missing data, many other aspects identified as crucial by other authors will typically not be mentioned. Regarding the problems we needed to solve, this often left us with more questions open than answered. For the future, a systematic, integrative, and comprehensive approach–specifically providing a hierarchy with regard to the severeness of one characteristic of missings in comparison with other characteristics–is very much needed.

6.2 DIRECTIONS FOR FUTURE RESEARCH

The current study examined the influence of emotional capabilities, SE cognitive ability, CIRME, and cognitive absorption on IR performance. Researchers investigating virtual worlds should consider looking at the impact of self-motivation and competition, as proposed by our secondary hypotheses, as well as at the impact of self-efficacy on performance (cf. also Joseph & Newman, 2010). Related thereto, we take from talking to virtual world users and from discussions among developers that self-motivation is being seen as an important driving force to accomplish extensive virtual world tasks, yet that requirements need to match users' current skill level to avoid frustration (Kiili, 2005). Furthermore, studies suggest that aspects like self-efficacy make important contributions to work-related performance (Judge et al., 2007), thus the question arises how important this factor may be in relation to virtual world tasks and IR performance. Interestingly, how individuals adapt their goal orientation, for instance in the face of failure, seems to depend on their cognitive ability. Thus whether an individual is learning-oriented or rather performance-oriented may lead to different effects on that individual's learning outcomes of self-efficacy, performance, and knowledge (B. S. Bell & Kozlowski, 2002). The path to understanding the relationship between competition and performance, however, has proved to be fraught with pitfalls, and the need for focusing on the combination of personality traits and situational variables in that context has recently been stressed in literature (Murayama & Elliot, 2012; Johnson et al., 2012). It is also very important to note that when implementing measures of emotional capabilities, the question whether one is dealing with men or with women significantly affects the outcome of a study (Brackett et al., 2006); hence if the proportion of women in our study had been higher, we might have seen different results.

With regard to serious applications of virtual worlds, the words of F. D. Davis (1989) are still valid: "Although difficulty of use can discourage adoption of an otherwise useful system, no amount of ease of use can compensate for a system that does not perform a useful function" (pp. 333–334). It is known from previous studies that "technology must be a good fit with the tasks it supports" and that "both utilization and user attitudes about the technology lead to individual performance impacts" (Goodhue & Thompson, 1995, p. 213). However, the perceived usefulness of virtual worlds may depend on the cognitive style–innovative or adaptive–of their users (Chakraborty, Hu, & Cui, 2008). Studies with telecommuters (who work from remote at home rather than in an on-site office) showed that the level of IS technologies available influence their individual performance and that the level of communication technologies available not only impacts performance, but also productivity and satisfaction (Bélanger et al., 2001). Also, performance and outcome related to multitasking have been found to depend on the level of multitasking and individual needs on diverse information (Aral et al., 2012). A large body of research has previously investigated knowledge adoption, management, and sharing practices in the past (e. g., Constant, Kiesler, & Sproull, 1994; Ngwenyama & Lee, 1997; Sussman & Sproull, 1999; Butler, 2001; K. H. Lim et al., 2000; Miranda & Saunders, 2003; Watts Sussman & Schneier Siegal, 2003; Wasko & Faraj, 2005; R. Baskerville & Dulipovici, 2006; Hsu & Lin, 2008; Albors, Ramos, & Hervas, 2008; Durcikova, Fadel, Butler, & Galletta, 2011), and research has begun to apply a specific virtual world lens regarding these issues (Berente et al., 2011; Kohler et al., 2011; J. Mueller et al., 2011). However, we believe that this area needs further scrutiny to better understand the impact of using virtual worlds for nonleisure applications, particularly in nonvoluntary settings like work contexts. Of particular interest to virtual worlds are, in our view, findings which associate task-specific representations to a significant improvement of task performance. Researchers suggest that users should be enabled to create new representations as needed (Morris, Neuwirth, Regli, Chandhok, & Wenger, 1999); we believe that virtual worlds have the potential to supply the required level of flexibility and thus to provide exciting opportunities in this regard.

According to collaborative media research, perceptions of the information culture, attitudes regarding information ownership, and the propensity among employees to share information are crucial conditions that shape a working environment and represent important factors for organizations to consider (cf. Jarvenpaa & Staples, 2000), as they ultimately influence work performance. Mennecke et al. (2008) have already identified 10 excellent research questions "worth considering" (pp. 382–383) in virtual worlds, most of them with reference to essential work-related topics, along with a roadmap for further research. There are also practical examples of companies attempting to provide guidance for their staff on how to deal with virtual worlds' unique characteristics and how to meet the challenges that arise when incorporating them into real-life business. IBM, for example, has set up specific virtual world procedures in alignment with the company's business code of conduct, dubbed the virtual worlds guidelines for IBM employees by the interested public. Questions that are being dealt with include:[1] How should an avatar look like when interacting with clients? How should a representative of the company behave in a virtual place which is similarly public as a hotel lobby or an airport? What rules apply in terms of intellectual property, virtual identity, and privacy? Future work is yet necessary on how traditional views on concepts like "copyright, code, creativity, and community–as well as a complementary component: the contract" (Roquilly, 2011, p. 667)–all of which are closely related to virtual worlds–will have to be adapted to the requirements imposed by actual virtual world practices.

1 Cited with IBM's permission.

Our last recommendation refers to the fact that social influence or social outcomes are often implemented as predictors of use intentions of virtual worlds. It seems that users of virtual worlds may instead regard the other participants as something that is embedded in the interaction and inherent to the medium (Mantymaki & Riemer, 2011), a perspective which better accounts for the characteristics of virtual worlds and may therefore be more suitable when investigating them.

We would like to conclude by expressing our wish for more cooperation as well as a better exchange of information and best-practice examples between the various scientific disciplines. Our initial literature review and experiences with studies from neighboring fields revealed an enormous knowledge base, for instance related to multiple aspects which may relate to virtual worlds. Computers are affecting almost every aspect of our lives, and it is not unlikely that one day the same will be true of virtual worlds. Arguably, virtual worlds provide (almost) CO_2-neutral means of travel and communication, which proves their potential to reduce our energy needs and thus solve urgent environmental challenges. Also, they may not only help to diffuse information, but as an excellent means for education and training, they may also support the interpretation, understanding, and application of this information, ultimately leading to knowledge creation and more self-determination for the people of this planet. We are certain that if scientists of various areas of psychology, computer science, HCI, engineering, societal sciences, and many others came together and jointly worked on the hard problems of our times, amazing things were to come.

We are aware of the fact that at best, research can only reflect recent state-of-the-art with regard to research design, methodology, measurement, and interpretation. It is limited by the precision of the measuring instruments currently available, and conclusions are biased by common perspectives and research paradigms. Not too long ago, Freedman (1991) warned that “statistical technique can seldom be an adequate substitute for good design, relevant data, and testing predictions against reality in a variety of settings” (p. 291), while MacCallum (2003) reminded us that all models “are wrong to some degree and are thus implausible if taken literally” (p. 113); another statement giving food for thought was given by A. S. Lee and Baskerville (2003): “As a consequence of Hume’s truism, a theory may never be generalized to a setting where it has not yet been empirically tested and confirmed” (p. 241). Others have demanded to exercise prudence when estimating intervention effects and have cautioned to speak of causal explanations when interpreting research findings (K. A. Markus, 2010; Mithas & Krishnan, 2009). Finally, Wilcox (1998) asked the provocative question “How many discoveries have been lost by ignoring modern statistical methods?” (p. 300). However, we believe that, by building on prior research and carefully replicating existing studies in order to reduce effects of capitalization on chance (Patil et al., 2008), we followed the advise to look for "significant sameness (...) rather than significant differences" (Hubbard, 1995, p. 1098)–true to the motto "standing on the shoulders of giants".

A

ITEMS AND SCALES

A.1 WORDINGS, ITEM SOURCES, AND SCALE FORMATION

Tables A.1-A.15 display the wordings of our items as well as of the items shared with the operator, grouped according to their construct or function (in German: all; in English: where applicable, the original items are taken from the sources indicated). For items and scales that were not newly developed, sources and the original wordings (including original numbering, punctuation, capitalization and other details) are given. Due to space constraints we start with the CIRME table first and then proceed in the usual order of constructs.

A.1.1 *Primary Constructs*

Table A.1: Childhood Inter-Reality Media Experience

Item	**Betrachte bitte den Zeitraum von Deinem 10. bis zum 14. Lebensjahr:** *Im Vergleich zu meinen Klassenkameraden habe ich...*
1) Watching IR action passively	[-3] sehr viel seltener – sehr viel öfter [+3]; keine Antwort
watch10_14_1	...anderen beim Computer- oder Konsolenspielen zugesehen ohne selber mitzuspielen.
watch10_14_2	...Spaß dabei gehabt, anderen beim Computer oder Konsolenspielen zuzusehen ohne selber mitzuspielen.
watch10_14_3	...die Aufforderung erhalten, anderen beim Computer- oder Konsolenspielen zuzusehen, ohne selber mitspielen zu können.
2) Experience IR on one's own & alone	[-3] sehr viel seltener – sehr viel öfter [+3]; keine Antwort
alone10_14_1	...alleine (d.h. ohne Freunde, Verwandte etc.) Computer- oder Konsolenspiele gespielt.
alone10_14_2	...mich alleine (d.h. ohne Freunde, Verwandte etc.) mit Computer- oder Konsolenspielen in meiner Freizeit beschäftigt.
alone10_14_3	...selbst bei schönem Wetter alleine (d.h. ohne Freunde, Verwandte etc.) Computer- oder Konsolenspielen gespielt, statt mit anderen Kindern im Freien.
3) Experience IR with others in physical proximity	[-3] sehr viel seltener – sehr viel öfter [+3]; keine Antwort
wothers10_14_1	...zusammen mit anderen vor dem Bildschirm gesessen, um Computer- oder Konsolenspiele zu spielen.

Continued on next page

Table A.1 Childhood IR Media Experience – *Continued*

wothers10_14_2	...Freunde besucht oder wurde besucht, um zusammen Computer- oder Konsolenspiele zu spielen.
wothers10_14_3	...zusammen mit Freunden im gleichen Raum Computer- oder Konsolenspiele gespielt.
4) Experience IR with others via network	[-3] sehr viel seltener – sehr viel öfter [+3]; keine Antwort
internwo10_14_1	...die Möglichkeit erhalten, über das Internet zusammen mit anderen zu spielen.
internwo10_14_2	...Computer- oder Konsolenspiele über das Internet zusammen mit anderen gespielt.
internwo10_14_3	...mich mit anderen zum Computer- oder Konsolenspielen via Internet verabredet.

Table A.2: Inter-Reality Emotional Capabilities

	Wie stark stimmst Du den folgenden Aussagen zu? *Wenn ich eSports ausübe...*		
Item		Source	Original item
1a) Perceive own emotions	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
perc_s_1	...kann ich Gefühle in mir wahrnehmen, sobald sie anfangen, sich zu entwickeln.	TEIQUE 1.5	126. I can identify an emotion from the moment it starts to develop in me.
perc_s_2	Bevor ich in einem Spiel sehr wütend werde, bemerke ich das oft schon lange vorher.	Adapted from TEIQUE 1.5	126. I can identify an emotion from the moment it starts to develop in me.
perc_s_3	Wenn ich einmal in einem Spiel sehr glücklich werde, kann ich das oft schon lange vorher vorhersehen.	"	126. I can identify an emotion from the moment it starts to develop in me.
perc_s_4	Bevor ich einem Spiel aggressiv werde, staut sich dieses Gefühl schon lange vorher in mir auf.	"	126. I can identify an emotion from the moment it starts to develop in me.
1b) Perceive others' emotions	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
perc_o_1	...kann ich beobachten, wenn ein Spieler im Spiel wütend ist.	Own item; observing and recognizing emotions in others	
perc_o_2	...kann ich beobachten, wenn ein Spieler glücklich ist.	"	
perc_o_3	...kann ich beobachten, wenn ein Spieler ängstlich ist.	"	

Continued on next page

Table A.2 IR Emotional Capabilities – *Continued*

perc_o_4	...weiß ich durch das Verhalten meiner Freunde im Spiel immer, welche Emotionen und Gefühle sie empfinden.	Wong and Law (2002); Others' emotion appraisal (OEA)	5. I always know my friends' emotions from their behavior.
perc_o_5	...bin ich ein guter Beobachter der Emotionen und Gefühle der anderen im Spiel.	"	6. I am a good observer of others' emotions.
perc_o_6	...erkenne ich durch das Verhalten der anderen im Spiel immer deren Emotionen und Gefühle.	Adapted from Wong and Law (2002); Others' emotion appraisal (OEA)	5. I always know my friends' emotions from their behavior.
perc_o_7	...kann ich die Gefühle der meisten Spieler lesen wie ein offenes Buch.	TEIQUE 1.5	17. I'm able to "read" most people's feelings like an open book.
perc_o_8	...nehme ich die Gefühle und Emotionen der anderen Spieler wahr.	"	2. Generally, I don't take notice of other people's emotions.
2a) Understand own emotions	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
underst_s_1	Ich finde leicht die richtigen Worte, um die Gefühle zu beschreiben, die ich beim e-Sport erlebe.	Adapted from TEIQUE	97. It is easy for me to find the right words to describe my feelings.
underst_s_2	...weiß ich meistens ganz genau, warum ich mich wie fühle.	Wong and Law (2002); Self-emotion appraisal (SEA)	1. I have a good sense of why I have certain feelings most of the time.
underst_s_3	...habe ich ein gutes Verständnis meiner eigenen Emotionen.	"	2. I have good understanding of my own emotions.
2b) Understand others' emotions	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
underst_o_1	...und ein anderer Spieler im Spiel plötzlich überrascht ist, kann ich mir gut vorstellen, woran das liegen könnte.	Own item; understanding emotions in others	
underst_o_2	...und ein anderer Spieler im Spiel plötzlich dankbar ist, kann ich mir gut vorstellen, woran das liegen könnte.	"	

Continued on next page

Table A.2 IR Emotional Capabilities – *Continued*

underst_o_3	...habe ich ein gutes Verständnis der Emotionen und Gefühle der anderen Spieler im Spiel.	Wong and Law (2002); Others' emotion appraisal (OEA)	8. I have good understanding of the emotions of people around me.
underst_o_4	...kann ich mich in die Gefühle und Emotionen der anderen Spieler im Spiel einfühlen.	Adapted from TEIQUE-SF	17. I'm normally able to "get into someone's shoes" and experience their emotions.
underst_o_5	...ist es kein Problem für mich, die Bedürfnisse und Verlangen der anderen Spieler im Spiel nachzuvollziehen.	TEIQUE 1.5	17. Understanding the needs and desires of others is not a problem for me.
underst_o_6	...fällt es mir oft schwer, die Dinge aus der Sicht eines anderen Spielers zu sehen. (R)*	"	42. I often find it difficult to see things from another person's viewpoint.
underst_o_7	...finde ich es schwer, die Emotionen anderer Spieler im Spiel zu verstehen. (R)*	"	49. I find it difficult to sympathize with other people's plights.
underst_o_8	...kann ich das Spiel leicht durch die Brille eines anderen Spielers sehen.	Adapted from TEIQUE 1.5	42. I often find it difficult to see things from another person's viewpoint.
3a) Manage own emotions	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
feel_man_1	...und unter Druck gesetzt werde, verliere ich leicht meine Beherrschung. (R)*	TEIQUE 1.5	136. When I'm under pressure, I tend to lose my cool.
feel_man_2	...erzählen mir andere, dass ich leicht gestresst werde. (R)*	"	139. Others tell me that I get stressed very easily.
feel_man_3	...habe ich meine Launen immer im Griff und handele auch in schwierigen Situation immer rational und besonnen.	Wong and Law (2002); Regulation of emotion (ROE)	13. I am able to control my temper and handle difficulties rationally.
feel_man_4	...und mich über etwas sehr ärgere, kann ich mich auch sehr schnell wieder beruhigen.	"	15. I can always calm down quickly when I am very angry.

Continued on next page

Table A.2 IR Emotional Capabilities – *Continued*

feel_man_5	...habe ich eine gute Kontrolle über meine Emotionen und Gefühle.	Wong and Law (2002); Regulation of emotion (ROE)	16. I have good control of my own emotions.
feel_man_6	...hat meine Stimmung nur geringe Auswirkungen darauf, wie ich an Probleme herangehe.	E. J. Austin et al. (2004)	4. My mood has little effect on how I deal with problems.
feel_man_7	...und vor ein Hindernis gerate, das meinen Erfolg gefährden könnte, erinnere ich mich daran, wie ich solche Situationen und damit verbundene Emotionen in der Vergangenheit erfolgreich gemeistert habe.	“	2. When I am faced with obstacles, I remember times when I faced similar obstacles and overcame them.
3b) Manage others' emotions	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
feel_o_1	...weiß ich normalerweise, was ich machen muss, um die Langeweile eines anderen Spielers in Aufregung zu verwandeln.	Tett et al. (2005); Regulation of emotion in others (RegOth)	Usually, I know what it takes to turn someone else's boredom into excitement.
feel_o_2	...und ich es wollte, wäre es einfach für mich, jemandem ein schlechtes Gefühl zu geben.	TEIQUE 1.5	60. If I wanted to, it would be easy for me to make someone feel bad.
feel_o_3	...scheine ich keine Macht über die Gefühle der anderen Spieler zu haben. (R)*	TEIQUE-SF	26. I don't seem to have any power at all over other people's feelings.
feel_o_4	...und ich es wollte, wäre es einfach für mich, jemanden sauer zu machen.	TEIQUE 1.5	86. If I wanted to, it would be easy for me to make someone angry.
feel_o_5	...kann ich normalerweise die Gefühle der anderen Spieler beeinflussen.	TEIQUE-SF	11. I'm usually able to influence the way other people feel.
feel_o_6	...weiß ich nicht, was ich machen muss, damit andere sich besser fühlen, wenn sie das brauchen. (R)*	TEIQUE 1.5	122. I don't know how to make others feel better when they need it.
feel_o_7	...und ich es wollte, wäre es einfach für mich, die Aufregung eines Spielers in Wut zu verwandeln.	Adapted from TEIQUE 1.5	86. If I wanted to, it would be easy for me to make someone angry.
feel_o_8	...und ich es wollte, wäre es einfach für mich, den Spaß eines Spielers in Frustration zu verwandeln.	“	86. If I wanted to, it would be easy for me to make someone angry.

Continued on next page

Table A.2 IR Emotional Capabilities – *Continued*

feel_o_9	...und ich es wollte, wäre es einfach für mich, die Zufriedenheit eines Spielers in Enttäuschung zu verwandeln.	Adapted from TEIQUE 1.5	86. If I wanted to, it would be easy for me to make someone angry.

* (R) denotes items that are reverse scored.

Table A.3: Self-Estimated Cognitive Ability

Item	**Wie stark stimmst Du den folgenden Aussagen zu?**	Source	Targeted facet
	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
clever_1	Mir sagen die Leute sehr oft, ich sei schlau.	Own item; compare S. J. Simon et al. (1996) & Ackerman and Wolman (2007)	Astuteness
clever_2	Sehr oft sind die Leute erstaunt darüber, wie schnell ich Dinge verstehe, die neu für mich sind.	“	Quick comprehension
clever_3	Mir fällt es sehr leicht, Dinge nachzumachen, die mir einer vormacht.	“	Quick learner
clever_4	Die Leute sagen, ich sei außergewöhnlich und außerordentlich intelligent.	Own item; compare S. J. Simon et al. (1996)	Above-average capacities
clever_5	Die Leute sagen, ich sei ein “sehr helles Köpfchen” mit guter Allgemeinbildung.	“	Educational background

Continued on next page

Table A.3 SE Cognitive Ability – *Continued*

clever_6	Die Leute sagen, in akademischen Dingen sei ich hochbegabt oder besonders talentiert.	Own item; compare S. J. Simon et al. (1996)	Intellectual giftedness
clever_7	In jedem Fach in der Schule gehörten meine Noten üblicherweise zu den besten.	"	Academic achievements

Table A.4: Cognitive Absorption

Item	**Wie stark stimmst Du den folgenden Aussagen zu?** *Wenn ich eSports ausübe...*	Source	Original item
1) Temporal dissociation	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
cogabs_1	...scheint die Zeit sehr schnell zu vergehen.	Agarwal and Karahanna (2000); Temporal Dissociation	TD1. Time appears to go by very quickly when I am using the Web.
cogabs_2	...vergesse ich manchmal die Zeit.	"	TD2. Sometimes I lose track of time when I am using the Web.
cogabs_3	...verbringe ich am Ende meistens mehr Zeit damit, als ich ursprünglich geplant hatte.	"	TD4. Most times when I get on to the Web, I end up spending more time that I had planned.
cogabs_8	...vergeht die Zeit wie im Flug.	"	TD3. Time flies when I am using the Web.

Continued on next page

Table A.4 Cognitive Absorption – *Continued*

2) Focused immersion	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
cogabs_4	...tauche ich in die Spiele ein.	Agarwal and Karahanna (2000); Focused Immersion	FI2. While using the Web, I am absorbed in what I am doing.
cogabs_5	...bin ich in die Aufgaben vertieft, die ich in den Spielen zu erledigen habe.	“	FI3. While on the Web, I am immersed in the task I am performing.
cogabs_6	...lasse ich mich leicht von anderen Dingen ablenken (z.B. von zwei Freunden, die sich neben mir über andere Themen unterhalten). (R)*	Agarwal and Karahanna (2000); Focused Immersion	FI1. While using the Web I am able to block out most other distractions.
cogabs_7	...gilt meine volle Aufmerksamkeit dem eSport.	“	FI5. While on the Web, my attention does not get diverted very easily.

* (R) denotes items that are reverse scored.

A.1.2 *Secondary Constructs*

Table A.5: Self-Motivational Traits

Item	**Bitte beantworte die folgenden Fragen!**	Source	Original item
1) Capability of self-motivation	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
motivation_1	Ich setze mir selber Ziele und versuche dann das Beste, um sie zu erreichen.	Wong and Law (2002); Use of emotion (UOE)	9. I always set goals for myself and then try my best to achieve them.
motivation_2	Ich sage mir immer wieder selbst, dass ich doch eine sehr fähige Persönlichkeit bin.	“	10. I always tell myself I am a competent person.
motivation_3	Ich bin eine sehr selbstmotivierte Person.	“	11. I am a self-motivated person.
motivation_4	Ich ermutige mich immer wieder, mein Bestes zu versuchen.	“	12. I would always encourage myself to try my best.
2) Intrinsic goal orientation	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
pers_stand_4	Es ist wichtig für mich, dass ich in allem was ich tue, durch und durch kompetent bin.	Frost et al. (1990)	6. It is important to me that I be thoroughly competent in everything I do.
pers_stand_5	Ich bin sehr gut darin, meine Bemühungen darauf zu fokussieren ein Ziel zu erreichen.	“	16. I am very good at focusing my efforts on attaining a goal.
pers_stand_6	Ich habe extrem hohe Ziele.	“	19. I have extremely high goals.

Table A.6: Competitiveness

Item	**Wie stark stimmst Du den folgenden Aussagen zu?** *Wenn ich eSports ausübe...*	Source	Original item
1) Seeking superiority	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
perf_orient_1	...bin ich der Beste im Spiel.	Yperen and Janssen (2002); Performance Orientation: “I feel most successful in my job when ...”	1. I am the best.
perf_orient_2	...sind andere nicht so gut wie ich.	“	2. others cannot do as well as me.
perf_orient_3	...bringe ich bessere Leistungen im Spiel als die anderen Spieler der Liga.	“	3. I perform better than my colleagues.
perf_orient_4	...vermasseln es andere Spieler, aber ich nicht.	“	5. others mess up and I do not.
perf_orient_5	...bekomme ich Dinge im Spiel hin, die andere nicht hinbekommen.	“	6. I accomplish something where others failed.
perf_orient_6	...bin ich einer der wenigen, der gewisse Dinge kennt oder eine spezielle Fähigkeit im Spiel hat.	“	7. I am the only one who knows about particular things or who has a particular skill.
perf_orient_7	...bin ich der effektivste Spieler.	Yperen and Janssen (2002); Performance Orientation: “I feel most successful in my job when ...”	8. I am clearly the most productive employee.
perf_orient_8	...kommen meine taktischen Fähigkeiten im Spiel sehr gut zur Geltung.	Own item; compare Yperen and Janssen (2002)	

Continued on next page

Table A.6 Competitiveness – *Continued*

2) Striving for perfection	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
perfect_1	...versuche ich, so perfekt wie möglich zu sein.	Stoeber et al. (2008); Striving for Perfection	1. I strive to be as perfect as possible.
perfect_2	...ist es wichtig für mich, in allem perfekt zu werden.	“	2. It is important to me to be perfect in everything I attempt.
perfect_3	...habe ich das Bedürfnis, immer perfekt zu sein.	“	3. I feel the need to be perfect.
perfect_4	...bin ich bei der Erreichung meiner Ziele ein Perfektionist.	“	4. I am a perfectionist as far as my targets are concerned.
perfect_5	...habe ich den Wunsch, alles perfekt zu machen.	“	5. I have the wish to do everything perfectly.
3) Personal standards in comparison with others	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort		
pers_stand_1	Andere Leute scheinen niedrigere Standards für sich selbst zu akzeptieren als ich.	Frost et al. (1990)	24. Other people seem to accept lower standards from themselves than I do.
pers_stand_2	Ich erwarte höhere Leistung in meinen täglichen Aufgaben als die meisten Leute.	Frost et al. (1990)	30. I expect higher performance in my daily tasks than most people.
pers_stand_3	Je mehr eSports ich betreibe, desto höher sind die Anforderungen, die ich im Vergleich zu anderen im Spiel an mich selbst stelle.	Adapted from Frost et al. (1990)	30. I expect higher performance in my daily tasks than most people.

Table A.7: Inter-Reality Enjoyment

Item	**Im Vergleich zu anderen Spielern bereitet mir eSports...**	Source	Original item
	[-3] sehr viel weniger – sehr viel mehr [+3]; keine Antwort		
fun_1	...Spaß.	Agarwal and Karahanna (2000); Heightened Enjoyment	HE1. I have fun interacting with the Web.
fun_2	...Vergnügen.	“	HE2. Using the Web provides me with a lot of enjoyment.
fun_3	...Aufregung.	Heijden (2004); Perceived Enjoyment	Exciting-dull
fun_4	...Genugtuung.	Adapted from W. W. Chin and Lee (2000); Overall Satisfaction Set 1	All things considered, I am: very dissatisfied / neither / very satisfied with using the system.

Table A.8: Inter-Reality Mediation

Item	**Betrachte bitte wieder den Zeitraum von Deinem 10. bis zum 14. Lebensjahr:** *Im Vergleich zu meinen Klassenkameraden hat mir…*	Source	Original item
	[-3] sehr viel seltener – sehr viel öfter [+3]; keine Antwort		
mediation_1	...jemand erklärt, warum manche Aktionen der anderen Spieler im Spiel erfolgreich sind.	Nikken (2003); Evaluative mediation	pointing to good things in a game

Continued on next page

Table A.8 IR Mediation – *Continued*

mediation_2	...jemand erklärt, warum manche Aktionen der anderen Spieler im Spiel nicht erfolgreich sind.	“	pointing to bad things in a game
mediation_3	...jemand erklärt, warum ein anderer Spieler eine bestimmte Aktion im Computer- oder Konsolenspiel macht.	Nikken (2003); Evaluative mediation	explaining what happens in games
mediation_4	...jemand erklärt, was beim Computer- oder Konsolenspielen in den anderen Spielern vorgeht.	Valkenburg et al. (1999); Instructive Mediation: “How often do you...”	4. ...explain the motives of TV characters?
mediation_5	...jemand klargemacht, wie Spiel und Wirklichkeit zusammenhängen.	“	5. ...explain what something on TV really means?

Table A.9: Parental Control

Item	**Betrachte bitte wieder den Zeitraum von Deinem 10. bis zum 14. Lebensjahr:** *Im Vergleich zu meinen Klassenkameraden haben meine Eltern...*	Source	Original item
	[-3] sehr viel weniger genau – sehr viel genauer [+3]; keine Antwort		
rules_parents_1	...gewusst, ob und welche Computer- oder Konsolenspiele ich gerade gespielt habe.	Valcke et al. (2010); Supervision	3. Afterwards, I control what my child watched on the Internet.
	[-3] sehr viel weniger – sehr viel mehr [+3]; keine Antwort		
rules_parents_2	...feste Regeln fürs Computer- oder Konsolenspielen auferlegt (z.B. nach dem Essen, nach den Hausaufgaben, spätabends etc.).	Valkenburg et al. (1999); Restrictive Mediation: “How often do you...”	7. ...set specific viewing hours for your child?

Continued on next page

Table A.9 Parental Control – *Continued*

	[-3] sehr viel weniger erfolgreich – sehr viel erfolgreicher [+3]; keine Antwort		
rules_parents_3	...durchgesetzt, dass ich nicht mehr spiele, als sie das wollten.	Valcke et al. (2010); Internet usage rules	8. I limit the time my child is allowed in the Internet (e. g., only one hour a day).
rules_parents_4	...durchgesetzt, welche Computer- oder Konsolenspielen ich spielen durfte.	Valcke et al. (2010); Stopping Internet usage	5. I stop my child when he/she visits a less suitable website.
	[-3] sehr viel seltener – sehr viel öfter [+3]; keine Antwort		
rules_parents_5	...feste Zeiten zum Computer- oder Konsolenspielen festgesetzt.	Valcke et al. (2010); Internet usage rules	7. I only allow my child to surf the Internet at specific days and times (e.g., only Wednesday afternoon).

A.1.3 *Control Items*

Table A.10: Controls Childhood

Item	
played10_14	Hast Du im Alter von 10-14 regelmäßig (d.h. mindestens alle 2-3 Wochen) anderen beim Computer- oder Konsolenspielen zugesehen oder selbst gespielt?
age_acc_console	Ab welchem Alter hattest Du allgemein Zugang (zu Hause, über Freunde, Verwandte, etc.) zu Computer- oder Konsolenspielen (Atari, Playstation, Nintendo, Xbox, etc.)?
network10_14	Hast Du im Alter von 10-14 regelmäßig (d.h. mindestens alle 2-3 Wochen) Computer- oder Konsolenspielen speziell im NETZWERK (LAN, Internet, etc.) gespielt?
age_acc_network	Ab welchem Alter hattest Du Zugang (zu Hause, über Freunde, Verwandte, etc.) zu Computer- oder Konsolenspielen, die speziell im NETZWERK (LAN, Internet, etc.) gespielt werden?
parents_lan_event	Haben Deine Eltern (oder Erziehungsberechtigten) Eltern-LAN Events besucht oder genutzt, um sich davon zu überzeugen, dass die ESL eine gute Sache ist?

A.1.4 *Preferences, Motives, Suggestions*

Table A.11: Gaming Preferences

Item	
num_games	Wie viele verschiedene Computer- oder Konsolenspiele übst Du als eSport aus?
favourite_game	Welches Computer- oder Konsolenspiel ist Dein Lieblingsspiel?
how_oft_new	Wie oft testest Du neue eSports-Games aus?
gamemod	Welche Spielmodi bevorzugst Du?
how_play	Wie übst Du e-Sports überwiegend aus?
success_feel_obs	Wie wichtig ist es für Deinen Erfolg in der ESL, die Gefühle der anderen Spieler wahrnehmen zu können?
detect_cheats [[Free text]]	Kannst Du im Spiel Cheats erkennen, wenn ja, woran?
	Würdest Du Dir wünschen, dass es mehr Möglichkeiten gäbe, Gefühle im Spiel wahrzunehmen?
	[[Multiple dichotomy group]]
Item	
fav_type_gm_1	Ego-Shooter
fav_type_gm_2	Actionspiele
fav_type_gm_3	Strategiespiele
fav_type_gm_4	Team-Shooter
fav_type_gm_5	Rollenspiele
fav_type_gm_6	MMOGs

Continued on next page

Table A.11 Gaming Preferences – *Continued*

fav_type_gm_7	Adventures
fav_type_gm_8	Gesellschaftsspiele
fav_type_gm_9	Jump'n'Run
fav_type_gm_10	Sportspiele
fav_type_gm_11	Simulationen
fav_type_gm_12	Rennspiele
fav_type_gm_13	DotA-Klone (Action-RTS)
fav_type_gm_14	facebook-Social Games
fav_type_gm_other [[Free text]]	Sonstiges:
fav_type_gm_999	keine Antwort

Table A.12: Motives for Playing

Item	**Wie sehr stimmst Du mit den folgenden Aussagen überein?**
	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort
reas_play_1	Ich nehme den Wettbewerb in der ESL sehr ernst.
reas_play_2	Gewinne (Sach- und Geldpreise) sind für mich der entscheidende Grund an Ligen teilzunehmen.
reas_play_3	Spielspaß und gute Matches sind für mich wichtiger als Gewinne und Preisgelder.

Table A.13: Request for Suggestions

Item	**Würdest Du Dir wünschen, dass es mehr Möglichkeiten gäbe, Gefühle im Spiel wahrzunehmen?** *Bitte wähle alle Punkte aus, die zutreffen:* [[Multiple dichotomy group]]
wish_feel_obs_1	Nein
wish_feel_obs_2	ja, im Text Chat
wish_feel_obs_3	ja, VOIP (z.B. Ventrilo, Skype, Team Speak etc.)
wish_feel_obs_4	ja, Video-Bild
wish_feel_obs_5	Emoticons an der Spielfigur
wish_feel_obs_999	keine Antwort
Item	**Wie sehr stimmst Du mit den folgenden Aussagen überein?**
	[-3] stimme gar nicht zu – stimme absolut zu [+3]; keine Antwort
ageclasses_yn	Sollte die ESL Altersklassen einführen, um Wettbewerbsgleichheit zu schaffen (etwa so wie beim Fußball die C-, B und A-Jugend oder U-21)?

Continued on next page

Table A.13 Request for Suggestions – *Continued*

bootcamp_yn	Sollte die ESL spezielle Boot Camps (Trainingslager) für junge Spieler organisieren, die nicht in Clans organisiert sind?
interrpt_gm_yn	Sollte die ESL es ermöglichen, bei beidseitigem Einvernehmen Spielstände zu speichern und damit die Möglichkeit bieten, laufende Spiele zu unterbrechen?
break_yn	Sollte die ESL pro Spiel jedem Team die Gelegenheit geben, das Spiel für maximal 5 Minuten zu unterbrechen (ähnlich wie eine Auszeit beim Basketball)?
gm_genre_offer	Welches Spiele-Genre soll die ESL in Zukunft verstärkt anbieten?

A.1.5 *Demographic Items*

Table A.14: Demographics

Item	
age [[Forced item, predetermined categories]]	Wie alt bist Du?
dem_1 [[Free text]]	Wie alt bist Du?
dem_2	Du bist...[m/w]?
dem_3	Dein Familienstand?
dem_4	Welcher Tätigkeit gehst Du zur Zeit nach?
dem_5	Wieviele Einwohner hat der Ort, in dem Du lebst?
dem_6	In welchem Bundesland lebst Du?
dem_8	Nenne uns Deinen höchsten bisher erreichten Bildungsabschluss.
dem_9	Welches ist der höchste Bildungsabschluss, den einer Deiner Erziehungsberechtigten erreicht hat?
dem_10	Ist Deutsch die Sprache, die Du zu Hause am meisten sprichst?
dem_11	In welchem Land wurdest Du geboren?
dem_12	In welchem Land lebst Du?

A.1.6 *Self-Disclosure and Comment*

Table A.15: Self-Disclosure

Item	
	Was trifft eher auf Dich zu?
answers_caref	Ich habe den Fragebogen nur schnell durchgeklickt, um einen Preis zu gewinnen. – Ich habe alle Fragen sorgfältig beantwortet; keine Antwort
	Diese Umfrage war...
survey_length	...viel zu lang – ...viel zu kurz

Continued on next page

Table A.15 Self-Disclosure – *Continued*

comment [[Free text]]	Hier kannst Du noch einen Kommentar oder eine Anregung los werden:

A.2 ITEMS IN ORDER OF APPEARANCE

The following table displays a list of variable names representing all items we sampled via the survey, in the same order as the corresponding items appeared in the survey. However, it does not include the operator's and system-generated items (e. g., like player IDs, see Section 3.2).

Table A.16: List of Variables in Order of Appearance

No.	Item	Described in
1	ageclasses_yn	Table A.13, Request for Suggestions
2	age	Table A.14, Demographics
3	num_games	Table A.11, Gaming Preferences
4	favourite_game	" , "
5	played10_14	Table A.10, Controls Childhood
6	age_acc_console	" , "
7	watch10_14_1	Table A.1, Childhood IR Media Experience
8	watch10_14_2	" , "
9	watch10_14_3	" , "
10	alone10_14_1	" , "
11	alone10_14_2	" , "
12	alone10_14_3	" , "
13	wothers10_14_1	" , "
14	wothers10_14_2	" , "
15	wothers10_14_3	" , "
16	network10_14	Table A.10, Controls Childhood
17	age_acc_network	" , "
18	internwo10_14_1	Table A.1, Childhood IR Media Experience
19	internwo10_14_2	" , "
20	internwo10_14_3	" , "
21	parents_lan_event	Table A.10, Controls Childhood
22	rules_parents_1	Table A.9, Parental Control
23	rules_parents_2	" , "
24	rules_parents_3	" , "
25	rules_parents_4	" , "
26	rules_parents_5	" , "
27	perfect_1	Table A.6, Competitiveness
28	perfect_2	" , "
29	perfect_3	" , "
30	perfect_4	" , "
31	perfect_5	" , "
32	clever_1	Table A.3, SE Cognitive Ability
33	clever_2	" , "
34	clever_3	" , "

Continued on next page

Table A.16 Variable List – *Continued*

No.	Item	Described in
35	pers_stand_1	Table A.6, Competitiveness
36	pers_stand_2	" , "
37	pers_stand_3	" , "
38	pers_stand_4	Table A.5, Self-Motivational Traits
39	pers_stand_5	" , "
40	pers_stand_6	" , "
41	clever_4	Table A.3, SE Cognitive Ability
42	clever_5	" , "
43	clever_6	" , "
44	clever_7	" , "
45	feel_o_1	Table A.2, IR Emotional Capabilities
46	feel_o_2	" , "
47	feel_o_3	" , "
48	feel_o_4	" , "
49	feel_o_5	" , "
50	feel_o_6	" , "
51	feel_o_7	" , "
52	feel_o_8	" , "
53	feel_o_9	" , "
54	success_feel_obs	Table A.11, Gaming Preferences
55	perc_o_1	Table A.2, IR Emotional Capabilities
56	perc_o_2	" , "
57	perc_o_3	" , "
58	perc_o_4	" , "
59	perc_o_5	" , "
60	perc_o_6	" , "
61	perc_o_7	" , "
62	perc_o_8	" , "
63	underst_o_1	" , "
64	underst_o_2	" , "
65	underst_o_3	" , "
66	underst_o_4	" , "
67	underst_o_5	" , "
68	underst_o_6	" , "
69	underst_o_7	" , "
70	underst_o_8	" , "
71	wish_feel_obs_1	Table A.13, Request for Suggestions
72	wish_feel_obs_2	" , "
73	wish_feel_obs_3	" , "
74	wish_feel_obs_4	" , "
75	wish_feel_obs_5	" , "
76	wish_feel_obs_999	" , "
77	bootcamp_yn	" , "
78	mediation_1	Table A.8, IR Mediation
79	mediation_2	" , "
80	mediation_3	" , "
81	mediation_4	" , "

Continued on next page

Table A.16 Variable List – *Continued*

No.	Item	Described in
82	mediation_5	Table A.8, IR Mediation
83	interrpt_gm_yn	Table A.13, Request for Suggestions
84	cogabs_1	Table A.4, Cognitive Absorption
85	cogabs_2	" , "
86	cogabs_3	" , "
87	cogabs_4	" , "
88	cogabs_5	" , "
89	cogabs_6	" , "
90	cogabs_7	" , "
91	cogabs_8	" , "
92	feel_man_1	Table A.2, IR Emotional Capabilities
93	feel_man_2	" , "
94	feel_man_3	" , "
95	feel_man_4	" , "
96	feel_man_5	" , "
97	feel_man_6	" , "
98	feel_man_7	" , "
99	break_yn	Table A.13, Request for Suggestions
100	perc_s_1	Table A.2, IR Emotional Capabilities
101	underst_s_2	" , "
102	underst_s_3	" , "
103	perc_s_2	" , "
104	perc_s_3	" , "
105	perc_s_4	" , "
106	underst_s_1	" , "
107	detect_cheats	Table A.11, Gaming Preferences
108	fun_1	Table A.7, IR Enjoyment
109	fun_2	" , "
110	fun_3	" , "
111	fun_4	" , "
112	how_oft_new	Table A.11, Gaming Preferences
113	motivation_1	Table A.5, Self-Motivational Traits
114	motivation_2	" , "
115	motivation_3	" , "
116	motivation_4	" , "
117	gm_genre_offer	Table A.13, Request for Suggestions
118	perf_orient_1	Table A.6, Competitiveness
119	perf_orient_2	" , "
120	perf_orient_3	" , "
121	perf_orient_4	" , "
122	perf_orient_5	" , "
123	perf_orient_6	" , "
124	perf_orient_7	" , "
125	perf_orient_8	" , "
126	gamemod	Table A.11, Gaming Preferences
127	how_play	" , "
128	dem_1	Table A.14, Demographics

Continued on next page

Table A.16 Variable List – *Continued*

No.	Item	Described in	
129	dem_2	Table A.14,	Demographics
130	dem_3	“ ,	“
131	dem_4	“ ,	“
132	dem_5	“ ,	“
133	dem_6	“ ,	“
134	dem_8	“ ,	“
135	dem_9	“ ,	“
136	dem_10	“ ,	“
137	dem_11	“ ,	“
138	dem_12	“ ,	“
139	fav_type_gm_1	Table A.11,	Gaming Preferences
140	fav_type_gm_2	“ ,	“
141	fav_type_gm_3	“ ,	“
142	fav_type_gm_4	“ ,	“
143	fav_type_gm_5	“ ,	“
144	fav_type_gm_6	“ ,	“
145	fav_type_gm_7	“ ,	“
146	fav_type_gm_8	“ ,	“
147	fav_type_gm_9	“ ,	“
148	fav_type_gm_10	“ ,	“
149	fav_type_gm_11	“ ,	“
150	fav_type_gm_12	“ ,	“
151	fav_type_gm_13	“ ,	“
152	fav_type_gm_14	“ ,	“
153	fav_type_gm_other	“ ,	“
154	fav_type_gm_999	“ ,	“
155	reas_play_1	Table A.12,	Motives for Playing
156	reas_play_2	“ ,	“
157	reas_play_3	“ ,	“
158	answers_caref	Table A.15,	Self-Disclosure
159	survey_length	“ ,	“
160	comment	“ ,	“

B DISTRIBUTION EXAMINATION

B.1 SKEWNESS AND KURTOSIS

The following tables display the skewness and kurtosis values for the univariate distributions of the construct variables, segregated according to primary and secondary constructs. Variables appear in order of appearance in the survey; the player level variable *esllevel* can be found amongst the primary constructs.

Table B.1: Values for Skewness and Kurtosis of Primary Construct Variables After Merging (N = 5,588)

Item	Valid n	Skew	Kurtosis
watch10_14_1	3171	0.054	-0.653
watch10_14_2	3155	-0.077	-0.846
watch10_14_3	3099	0.534	-0.626
allone10_14_1	3161	-0.521	-0.365
allone10_14_2	3161	-0.424	-0.445
allone10_14_3	3164	0.153	-0.905
wothers10_14_1	3147	-0.611	0.046
wothers10_14_2	3136	-0.695	0.018
wothers10_14_3	3105	-0.724	0.035
internwo10_14_1	3087	-0.299	-0.920
internwo10_14_2	3083	-0.310	-1.000
internwo10_14_3	3068	-0.203	-1.079
clever_1	3513	-0.584	0.279
clever_2	3516	-0.684	0.208
clever_3	3537	-0.792	0.406
clever_4	3463	-0.199	-0.328
clever_5	3475	-0.465	-0.158
clever_6	3412	-0.087	-0.407
clever_7	3462	0.025	-0.760
feel_o_1	3108	-0.318	0.008
feel_o_2	3075	-0.018	-0.482
feel_o_3	3064	-0.091	-0.289
feel_o_4	3101	-0.416	-0.464
feel_o_5	3089	-0.260	-0.151
feel_o_6	3073	-0.250	-0.595
feel_o_7	3055	-0.184	-0.649
feel_o_8	3061	-0.066	-0.785
feel_o_9	3045	0.008	-0.765
perc_o_1	3076	-0.696	0.230
perc_o_2	3070	-0.592	0.113
perc_o_3	3052	-0.367	-0.447
perc_o_4	3052	-0.543	0.108
perc_o_5	3033	-0.459	-0.052
perc_o_6	3037	-0.331	-0.118

Continued on next page

Table B.1 Skewness and Kurtosis - PC – *Continued*

	Valid n	Skew	Kurtosis
perc_o_7	3027	-0.109	-0.482
perc_o_8	3027	-0.500	0.003
underst_o_1	2804	-0.442	0.227
underst_o_2	2817	-0.450	0.240
underst_o_3	2810	-0.363	0.244
underst_o_4	2806	-0.326	0.104
underst_o_5	2804	-0.388	0.212
underst_o_6	2815	-0.319	-0.441
underst_o_7	2804	-0.267	-0.415
underst_o_8	2764	-0.240	-0.095
cogabs_1	2839	-0.826	0.383
cogabs_2	2837	-0.664	-0.100
cogabs_3	2829	-0.470	-0.420
cogabs_4	2829	-0.475	-0.278
cogabs_5	2826	-0.563	-0.027
cogabs_6	2807	-0.038	-0.805
cogabs_7	2816	-0.385	-0.456
cogabs_8	2805	-0.680	-0.085
feel_man_1	2710	-0.220	-0.628
feel_man_2	2709	-0.193	-0.635
feel_man_3	2714	-0.220	-0.360
feel_man_4	2710	-0.459	-0.091
feel_man_5	2707	-0.366	-0.123
feel_man_6	2679	-0.225	-0.170
feel_man_7	2639	-0.378	-0.119
perc_s_1	2611	-0.447	0.237
perc_s_2	2596	-0.408	-0.200
perc_s_3	2586	-0.154	-0.129
perc_s_4	2557	-0.206	-0.590
underst_s_1	2561	-0.337	-0.045
underst_s_2	2657	-0.654	0.350
underst_s_3	2650	-0.642	0.428
esllevel	5588	3.323	20.557

Table B.2: Values for Skewness and Kurtosis of Secondary Construct Variables After Merging (N = 5,588)

Item	Valid n	Skew	Kurtosis
rules_parents_1	3050	-.432	-.750
rules_parents_2	3962	.030	-1.021
rules_parents_3	3791	.054	-.763
rules_parents_4	3764	.255	-.877
rules_parents_5	3801	.289	-.869
mediation_1	2064	-.287	-.574
mediation_2	2070	-.316	-.480
mediation_3	2073	-.387	-.340
mediation_4	2059	-.154	-.617

Continued on next page

Table B.2 Skewness and Kurtosis - SC – *Continued*

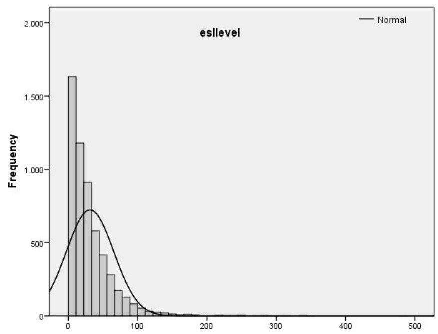
	Valid n	Skew	Kurtosis
mediation_5	2055	-.340	-.624
perf_orient_1	2583	-.236	-.165
perf_orient_2	2588	-.195	-.127
perf_orient_3	2570	-.090	-.040
perf_orient_4	2577	.005	-.093
perf_orient_5	2575	-.246	.120
perf_orient_6	2578	-.277	-.122
perf_orient_7	2569	-.072	-.076
perf_orient_8	2574	-.441	-.027
perfect_1	3793	-.747	-.070
perfect_2	3800	-.467	-.530
perfect_3	3794	-.405	-.665
perfect_4	3789	-.427	-.513
perfect_5	3775	-.688	-.329
pers_stand_1	3473	-.280	-.380
pers_stand_2	3493	-.242	-.475
pers_stand_3	3481	-.519	-.365
motivation_1	2649	-.715	.360
motivation_2	2618	-.544	-.040
motivation_3	2640	-.547	.027
motivation_4	2631	-.753	.407
pers_stand_4	3531	-.633	.092
pers_stand_5	3538	-.701	.229
pers_stand_6	3522	-.469	-.417
fun_1	2752	-.674	.211
fun_2	2745	-.591	.141
fun_3	2737	-.506	-.212
fun_4	2707	-.276	-.347

B.2 DISTRIBUTION DIAGRAMS

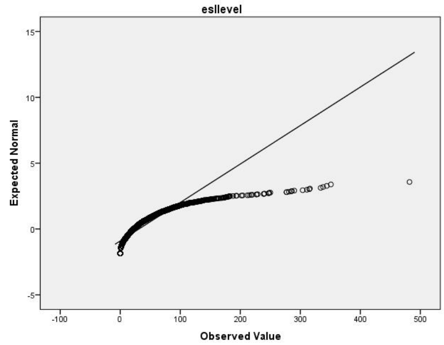
The following diagrams show the histograms with delineated normality plots (tagged with H) as well as normal the Q-Q plots diagrams (tagged with Q) for the primary construct variables, grouped by constructs (N = 5,588).

A Q-Q plot compares two distributions by plotting their quantiles against each other. Wilk and Gnanadesikan (1968) explain how to interpret of Q-Q plots like this: “If x and y are identically distributed variables, then the plot of x-quantiles versus y-quantiles will of course be a straight line configuration with slope 1, pointed towards the origin” (p. 5). Here, we compare the distribution of our data to the normal distribution (via a so-called normal Q-Q plot), so if the data follows the normal distribution, the points in the plot will approximately lie on the line $y = x$ (cf. also Kline, 2011, p. 209) .

Figure B.1: Diagrams for Performance

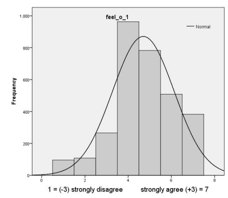


(a) H esllevel

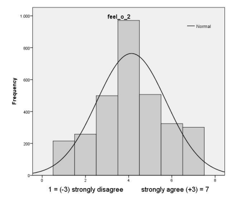


(b) Q esllevel

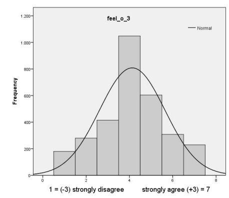
Figure B.2: Diagrams for IR Emotional Capabilities I



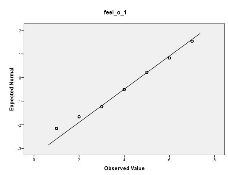
(a) H feel_o_1



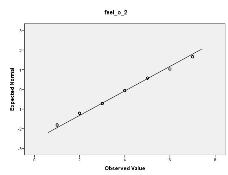
(b) H feel_o_2



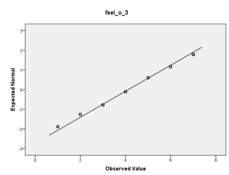
(c) H feel_o_3



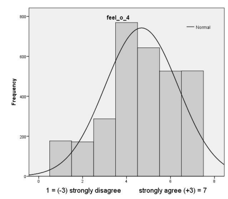
(d) Q feel_o_1



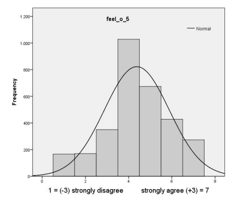
(e) Q feel_o_2



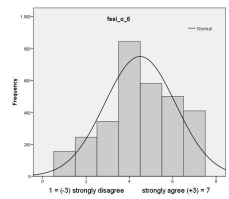
(f) Q feel_o_3



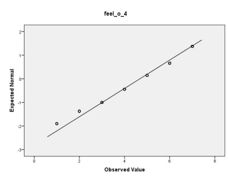
(g) H feel_o_4



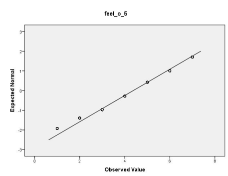
(h) H feel_o_5



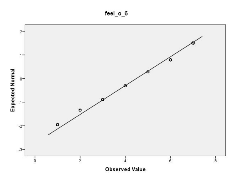
(i) H feel_o_6



(j) Q feel_o_4

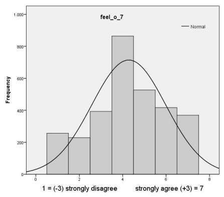


(k) Q feel_o_5

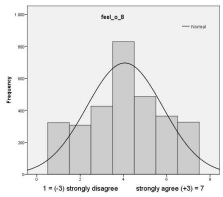


(l) Q feel_o_6

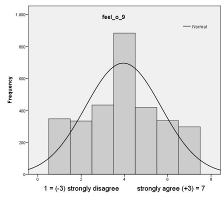
Figure B.3: Diagrams for IR Emotional Capabilities I *(continued)*



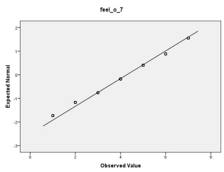
(m) H feel_o_7



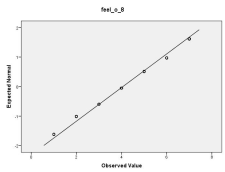
(n) H feel_o_8



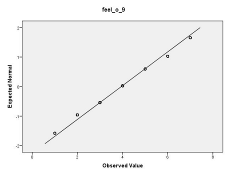
(o) H feel_o_9



(p) Q feel_o_7

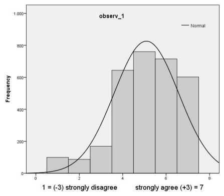


(q) Q feel_o_8

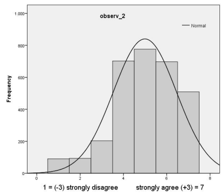


(r) Q feel_o_9

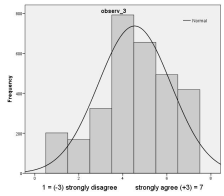
Figure B.4: Diagrams for IR Emotional Capabilities II



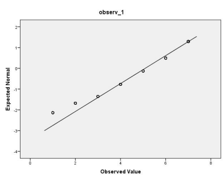
(a) H perc_o_1



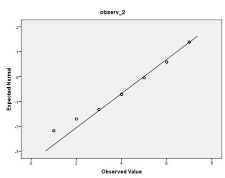
(b) H perc_o_2



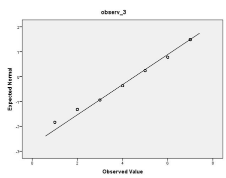
(c) H perc_o_3



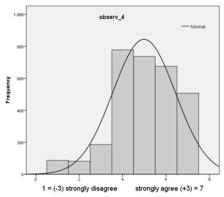
(d) Q perc_o_1



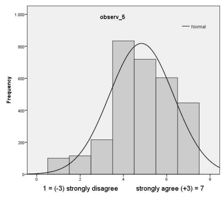
(e) Q perc_o_2



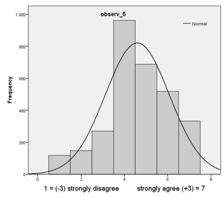
(f) Q perc_o_3



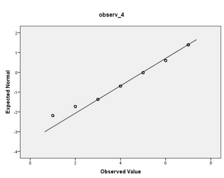
(g) H perc_o_4



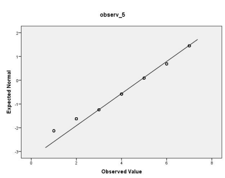
(h) H perc_o_5



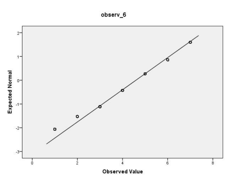
(i) H perc_o_6



(j) Q perc_o_4

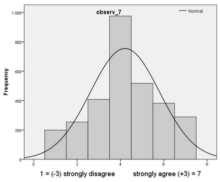


(k) Q perc_o_5

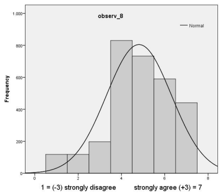


(l) Q perc_o_6

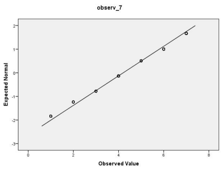
Figure B.5: Diagrams for IR Emotional Capabilities II *(continued)*



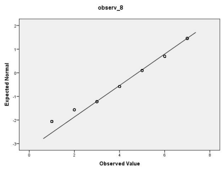
(m) H perc_o_7



(n) H perc_o_8

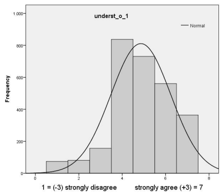


(o) Q perc_o_7

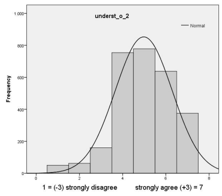


(p) Q perc_o_8

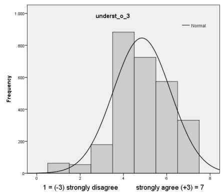
Figure B.6: Diagrams for IR Emotional Capabilities III



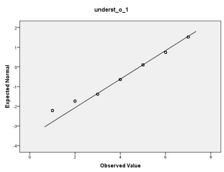
(a) H underst_o_1



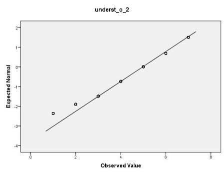
(b) H underst_o_2



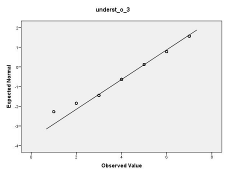
(c) H underst_o_3



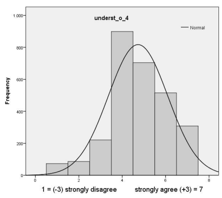
(d) Q underst_o_1



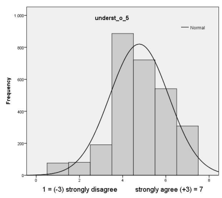
(e) Q underst_o_2



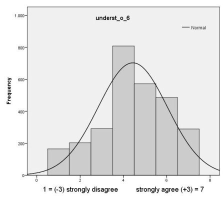
(f) Q underst_o_3



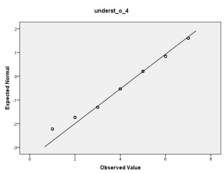
(g) H underst_o_4



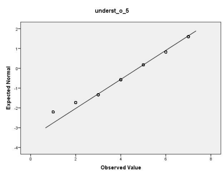
(h) H underst_o_5



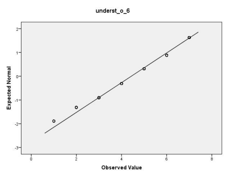
(i) H underst_o_6



(j) Q underst_o_4

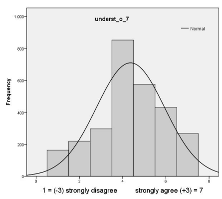


(k) Q underst_o_5

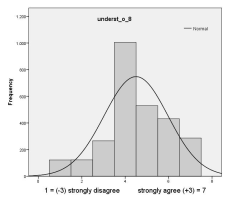


(l) Q underst_o_6

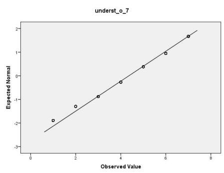
Figure B.7: Diagrams for IR Emotional Capabilities III *(continued)*



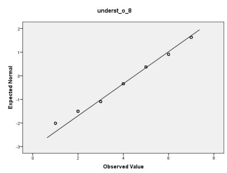
(m) H underst_o_7



(n) H underst_o_8

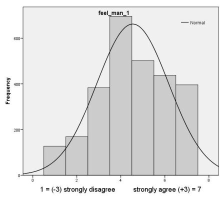


(o) Q underst_o_7

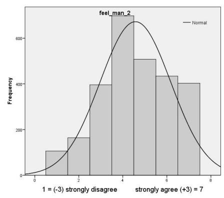


(p) Q underst_o_8

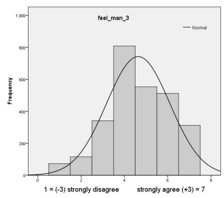
Figure B.8: Diagrams for IR Emotional Capabilities IV



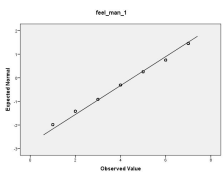
(a) H feel_man_1



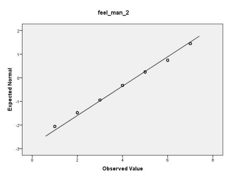
(b) H feel_man_2



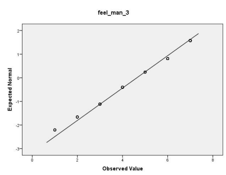
(c) H feel_man_3



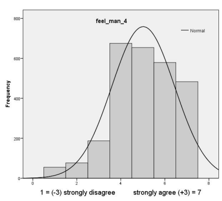
(d) Q feel_man_1



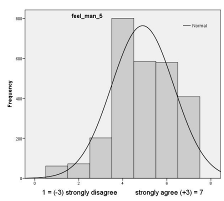
(e) Q feel_man_2



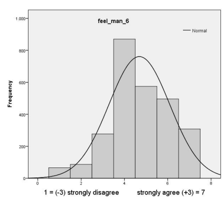
(f) Q feel_man_3



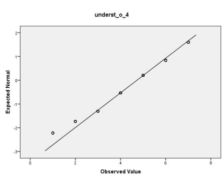
(g) H feel_man_4



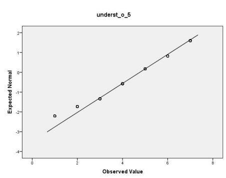
(h) H feel_man_5



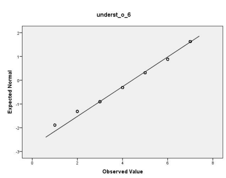
(i) H feel_man_6



(j) Q feel_man_4

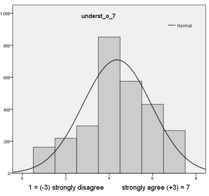


(k) Q feel_man_5

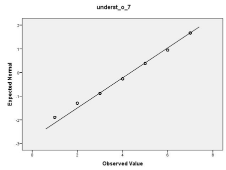


(l) Q feel_man_6

Figure B.9: Diagrams for IR Emotional Capabilities IV *(continued)*

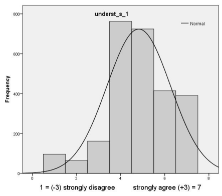


(m) H feel_man_7

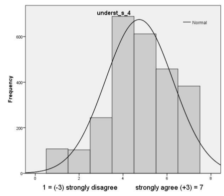


(n) Q feel_man_7

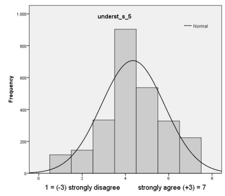
Figure B.10: Diagrams for IR Emotional Capabilities V



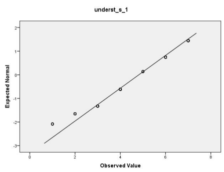
(a) H perc_s_1



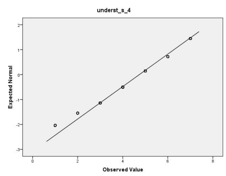
(b) H perc_s_2



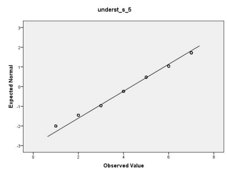
(c) H perc_s_3



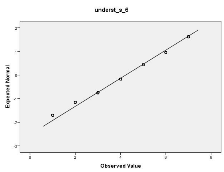
(d) Q perc_s_1



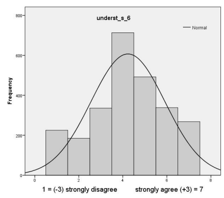
(e) Q perc_s_2



(f) Q perc_s_3

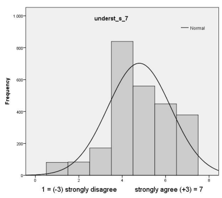


(g) Q perc_s_4

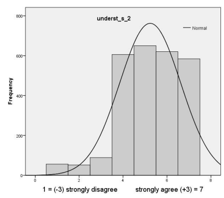


(h) H perc_s_4

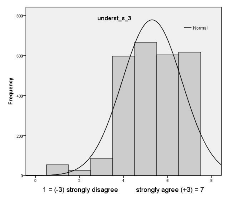
Figure B.11: Diagrams for IR Emotional Capabilities VI



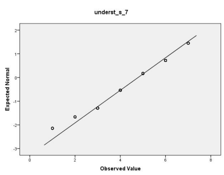
(a) H underst_s_1



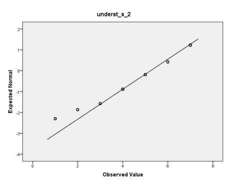
(b) H underst_s_2



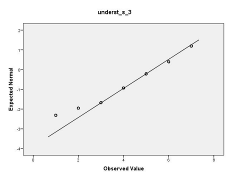
(c) H underst_s_3



(d) Q underst_s_1

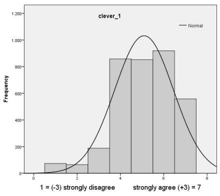


(e) Q underst_s_2

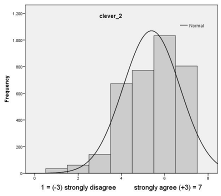


(f) Q underst_s_3

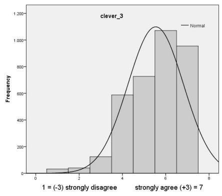
Figure B.12: Diagrams for SE Cognitive Ability



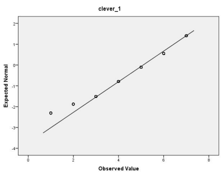
(a) H clever_1



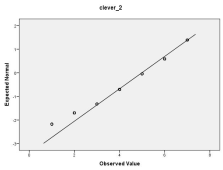
(b) H clever_2



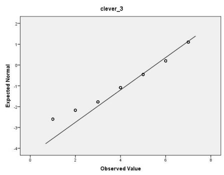
(c) H clever_3



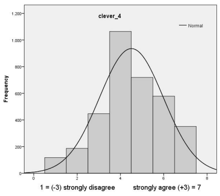
(d) Q clever_1



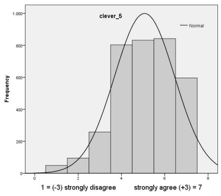
(e) Q clever_2



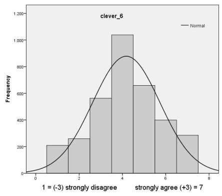
(f) Q clever_3



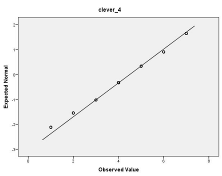
(g) H clever_4



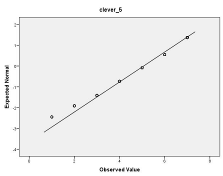
(h) H clever_5



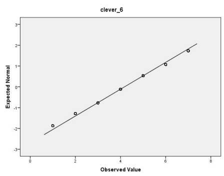
(i) H clever_6



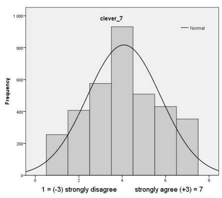
(j) Q clever_4



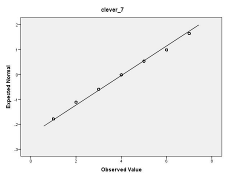
(k) Q clever_5



(l) Q clever_6

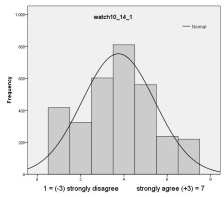


(m) H clever_7

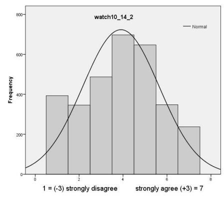


(n) Q clever_7

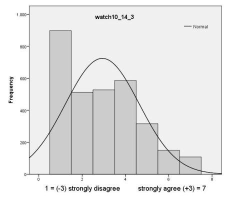
Figure B.13: Diagrams for CIRME



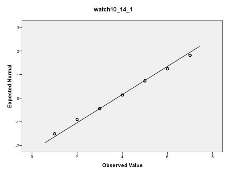
(a) H watch10__14__1



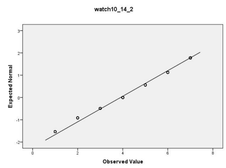
(b) H watch10__14__2



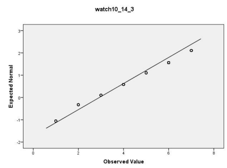
(c) H watch10__14__3



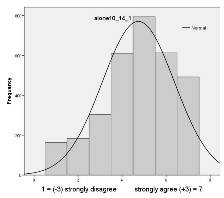
(d) Q watch10__14__1



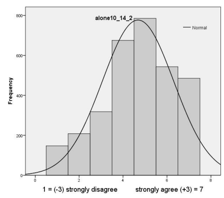
(e) Q watch10__14__2



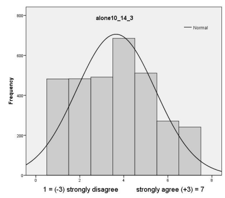
(f) Q watch10__14__3



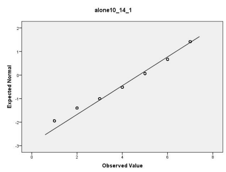
(g) H alone10__14__1



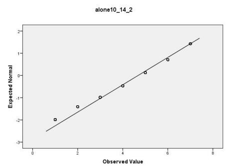
(h) H alone10__14__2



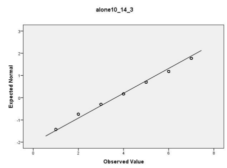
(i) H alone10__14__3



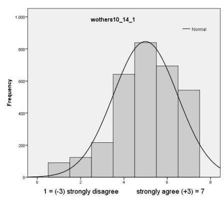
(j) Q alone10__14__1



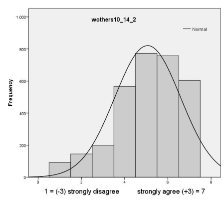
(k) Q alone10__14__2



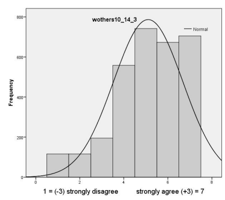
(l) Q alone10__14__3



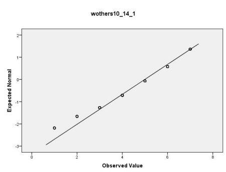
(m) H wothers10__14__1



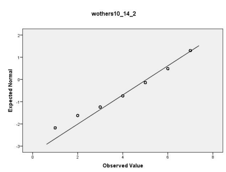
(n) H wothers10__14__2



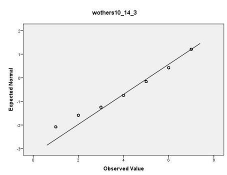
(o) H wothers10__14__3



(p) Q wothers10__14__1

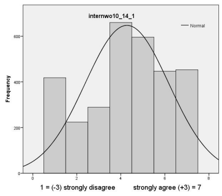


(q) Q wothers10__14__2

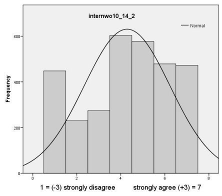


(r) Q wothers10__14__3

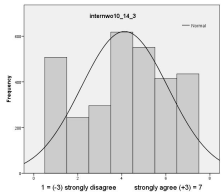
Figure B.14: Diagrams for CIRME *(continued)*



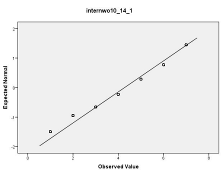
(s) H internwo10_14_1



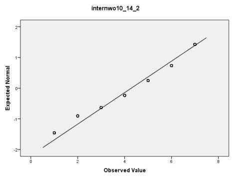
(t) H internwo10_14_2



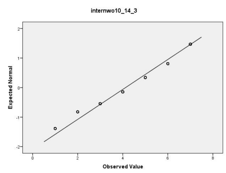
(u) H internwo10_14_3



(v) Q internwo10_14_1

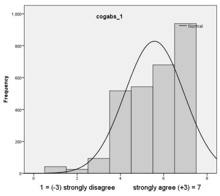


(w) Q internwo10_14_2

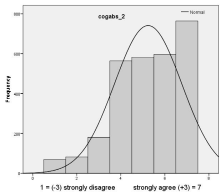


(x) Q internwo10_14_3

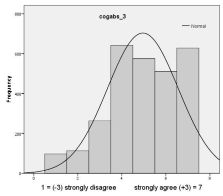
Figure B.15: Diagrams for Cognitive Absorption



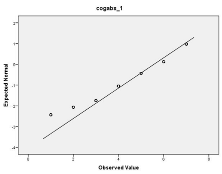
(a) H cogabs_1



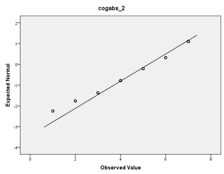
(b) H cogabs_2



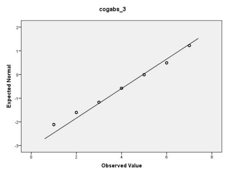
(c) H cogabs_3



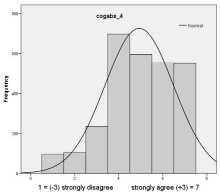
(d) Q cogabs_1



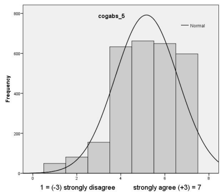
(e) Q cogabs_2



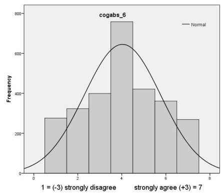
(f) Q cogabs_3



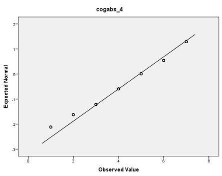
(g) H cogabs_4



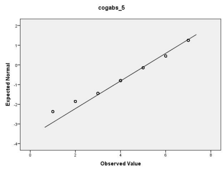
(h) H cogabs_5



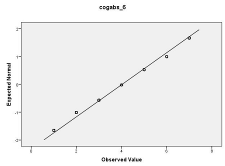
(i) H cogabs_6



(j) Q cogabs_4

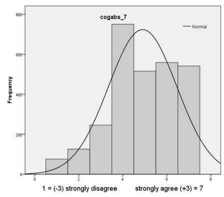


(k) Q cogabs_5

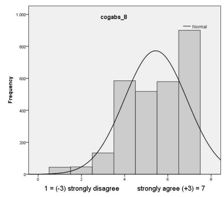


(l) Q cogabs_6

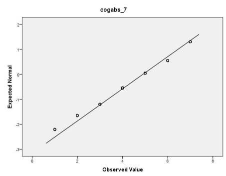
Figure B.16: Diagrams for Cognitive Absorption *(continued)*



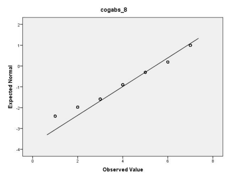
(m) H cogabs_7



(n) H cogabs_8



(o) Q cogabs_7



(p) Q cogabs_8

C MISSING DATA TREATMENT

C.1 EXTENT OF MISSINGS

The following tables show the total amount and percentages of missing values for all primary and secondary construct variables after merging the data of the unique group with selected entries of the nonunique group (displayed in order of appearance in the questionnaire).

Table C.1: Absolute Values and Percentages of Missings for Primary Construct Variables (N = 5,588)

	Valid		Missing	
Item	n	%	n	%
watch10_14_1	3171	56.7	2417	43.3
watch10_14_2	3155	56.5	2433	43.5
watch10_14_3	3099	55.5	2489	44.5
alone10_14_1	3161	56.6	2427	43.4
alone10_14_2	3161	56.6	2427	43.4
alone10_14_3	3164	56.6	2424	43.4
wothers10_14_1	3147	56.3	2441	43.7
wothers10_14_2	3136	56.1	2452	43.9
wothers10_14_3	3105	55.6	2483	44.4
internwo10_14_1	3087	55.2	2501	44.8
internwo10_14_2	3083	55.2	2505	44.8
internwo10_14_3	3068	54.9	2520	45.1
clever_1	3513	62.9	2075	37.1
clever_2	3516	62.9	2072	37.1
clever_3	3537	63.3	2051	36.7
clever_4	3463	62.0	2125	38.0
clever_5	3475	62.2	2113	37.8
clever_6	3412	61.1	2176	38.9
clever_7	3462	62.0	2126	38.0
feel_o_1	3108	55.6	2480	44.4
feel_o_2	3075	55.0	2513	45.0
feel_o_3	3064	54.8	2524	45.2
feel_o_4	3101	55.5	2487	44.5
feel_o_5	3089	55.3	2499	44.7
feel_o_6	3073	55.0	2515	45.0
feel_o_7	3055	54.7	2533	45.3
feel_o_8	3061	54.8	2527	45.2
feel_o_9	3045	54.5	2543	45.5
observ_1	3076	55.0	2512	45.0
observ_2	3070	54.9	2518	45.1
observ_3	3052	54.6	2536	45.4
observ_4	3052	54.6	2536	45.4

Continued on next page

Table C.1 Amount of Missings I – *Continued*

	Valid		Missing	
	n	%	n	%
observ_5	3033	54.3	2555	45.7
observ_6	3037	54.3	2551	45.7
observ_7	3027	54.2	2561	45.8
observ_8	3027	54.2	2561	45.8
underst_o_1	2804	50.2	2784	49.8
underst_o_2	2817	50.4	2771	49.6
underst_o_3	2810	50.3	2778	49.7
underst_o_4	2806	50.2	2782	49.8
underst_o_5	2804	50.2	2784	49.8
underst_o_6	2815	50.4	2773	49.6
underst_o_7	2804	50.2	2784	49.8
underst_o_8	2764	49.5	2824	50.5
cogabs_1	2839	50.8	2749	49.2
cogabs_2	2837	50.8	2751	49.2
cogabs_3	2829	50.6	2759	49.4
cogabs_4	2829	50.6	2759	49.4
cogabs_5	2826	50.6	2762	49.4
cogabs_6	2807	50.2	2781	49.8
cogabs_7	2816	50.4	2772	49.6
cogabs_8	2805	50.2	2783	49.8
feel_man_1	2710	48.5	2878	51.5
feel_man_2	2709	48.5	2879	51.5
feel_man_3	2714	48.6	2874	51.4
feel_man_4	2710	48.5	2878	51.5
feel_man_5	2707	48.4	2881	51.6
feel_man_6	2679	47.9	2909	52.1
feel_man_7	2639	47.2	2949	52.8
underst_s_1	2611	46.7	2977	53.3
underst_s_2	2657	47.5	2931	52.5
underst_s_3	2650	47.4	2938	52.6
underst_s_4	2596	46.5	2992	53.5
underst_s_5	2586	46.3	3002	53.7
underst_s_6	2557	45.8	3031	54.2
underst_s_7	2561	45.8	3027	54.2
esllevel	5588	100.0	0	0.0

Table C.2: Absolute Values and Percentages of Missings for Secondary Construct Variables (N = 5,588)

	Valid		Missing	
Item	n	%	n	%
rules_parents_1	3050	54.6	2538	45.4
rules_parents_2	3962	70.9	1626	29.1
rules_parents_3	3791	67.8	1797	32.2
rules_parents_4	3764	67.4	1824	32.6

Continued on next page

Table C.2 Amount of Missings II – *Continued*

	Valid			Missing	
	n	%		n	%
rules_parents_5	3801	68.0	1787	32.0	
mediation_1	2064	36.9		3524	63.1
mediation_2	2070	37.0		3518	63.0
mediation_3	2073	37.1		3515	62.9
mediation_4	2059	36.8		3529	63.2
mediation_5	2055	36.8		3533	63.2
perf_orient_1	2583	46.2		3005	53.8
perf_orient_2	2588	46.3		3000	53.7
perf_orient_3	2570	46.0		3018	54.0
perf_orient_4	2577	46.1		3011	53.9
perf_orient_5	2575	46.1		3013	53.9
perf_orient_6	2578	46.1		3010	53.9
perf_orient_7	2569	46.0		3019	54.0
perf_orient_8	2574	46.1		3014	53.9
perfect_1	3793	67.9		1795	32.1
perfect_2	3800	68.0		1788	32.0
perfect_3	3794	67.9		1794	32.1
perfect_4	3789	67.8		1799	32.2
perfect_5	3775	67.6		1813	32.4
pers_stand_1	3473	62.2		2115	37.8
pers_stand_2	3493	62.5		2095	37.5
pers_stand_3	3481	62.3		2107	37.7
motivation_1	2649	47.4		2939	52.6
motivation_2	2618	46.9		2970	53.1
motivation_3	2640	47.2		2948	52.8
motivation_4	2631	47.1		2957	52.9
pers_stand_4	3531	63.2		2057	36.8
pers_stand_5	3538	63.3		2050	36.7
pers_stand_6	3522	63.0		2066	37.0
fun_1	2752	49.2		2836	50.8
fun_2	2745	49.1		2843	50.9
fun_3	2737	49.0		2851	51.0
fun_4	2707	48.4		2881	51.6

C.2 MISSING PATTERNS

The next table shows the missing patterns of the primary construct variables as well as the number of their occurrence, “x” delineating missings; patterns with less than 1% cases (56 or fewer) are not displayed.

Table C.3: Patterns of Missings in Data File

No. of Cases	1467	80	90	70	483	156	96	388	97	1369
watch10_14_1					x		x			x
watch10_14_2					x		x			x
watch10_14_3					x		x			x
alone10_14_1					x		x			x
alone10_14_2					x		x			x
alone10_14_3					x		x			x
wothers10_14_1					x		x			x
wothers10_14_2					x		x			x
wothers10_14_3					x		x			x
internwo10_14_1					x		x		x	x
internwo10_14_2					x		x		x	x
internwo10_14_3					x		x		x	x
clever_1								x	x	x
clever_2								x	x	x
clever_3								x	x	x
clever_4								x	x	x
clever_5								x	x	x
clever_6								x	x	x
clever_7								x	x	x
feel_o_1						x	x	x	x	x
feel_o_2						x	x	x	x	x
feel_o_3						x	x	x	x	x
feel_o_4						x	x	x	x	x
feel_o_5						x	x	x	x	x
feel_o_6						x	x	x	x	x
feel_o_7						x	x	x	x	x
feel_o_8						x	x	x	x	x
feel_o_9						x	x	x	x	x
perc_o_1				x		x	x	x	x	x
perc_o_2				x		x	x	x	x	x
perc_o_3				x		x	x	x	x	x
perc_o_4				x		x	x	x	x	x
perc_o_5				x		x	x	x	x	x
perc_o_6				x		x	x	x	x	x
perc_o_7				x		x	x	x	x	x
perc_o_8				x		x	x	x	x	x
underst_o_1			x	x		x	x	x	x	x
underst_o_2			x	x		x	x	x	x	x
underst_o_3			x	x		x	x	x	x	x
underst_o_4			x	x		x	x	x	x	x
underst_o_5			x	x		x	x	x	x	x
underst_o_6			x	x		x	x	x	x	x
underst_o_7			x	x		x	x	x	x	x
underst_o_8			x	x		x	x	x	x	x
cogabs_1		x	x	x		x	x	x	x	x
cogabs_2		x	x	x		x	x	x	x	x
cogabs_3		x	x	x		x	x	x	x	x
cogabs_4		x	x	x		x	x	x	x	x
cogabs_5		x	x	x		x	x	x	x	x
cogabs_6		x	x	x		x	x	x	x	x
cogabs_7		x	x	x		x	x	x	x	x
cogabs_8		x	x	x		x	x	x	x	x
feel_man_1		x	x	x		x	x	x	x	x
feel_man_2		x	x	x		x	x	x	x	x
feel_man_3		x	x	x		x	x	x	x	x
feel_man_4		x	x	x		x	x	x	x	x
feel_man_5		x	x	x		x	x	x	x	x
feel_man_6		x	x	x		x	x	x	x	x
feel_man_7		x	x	x		x	x	x	x	x
perc_s_1		x	x	x		x	x	x	x	x
underst_s_2		x	x	x		x	x	x	x	x
underst_s_3		x	x	x		x	x	x	x	x
perc_s_2		x	x	x		x	x	x	x	x
perc_s_3		x	x	x		x	x	x	x	x
perc_s_4		x	x	x		x	x	x	x	x
underst_s_1		x	x	x		x	x	x	x	x

C.3 DATA BASE FOR EM AND FIML ESTIMATION

The next table shows the total amount and percentages of missing values for all primary construct variables after the elimination of 1,369 cases with no data on said variables (cf. Section 4.3). The patterns observed after elimination were the same as reported before (cf. Appendix C, Section C.2), yet naturally with the exception that the pattern with no data on the primary construct variables (Table C.3, furthest column to the right) was eliminated along with the removed cases.

Table C.4: Absolute Values and Percentages of Missings for Primary Construct Variables Prior to EM Estimation (N = 4,219)

Item	n	%
watch10_14_1	1048	24.84
watch10_14_2	1064	25.22
watch10_14_3	1120	26.55
alone10_14_1	1058	25.08
alone10_14_2	1058	25.08
alone10_14_3	1055	25.01
wothers10_14_1	1072	25.41
wothers10_14_2	1083	25.67
wothers10_14_3	1114	26.40
internwo10_14_1	1132	26.83
internwo10_14_2	1136	26.93
internwo10_14_3	1151	27.28
clever_1	706	16.73
clever_2	703	16.66
clever_3	682	16.16
clever_4	756	17.92
clever_5	744	17.63
clever_6	807	19.13
clever_7	757	17.94
feel_o_1	1111	26.33
feel_o_2	1144	27.12
feel_o_3	1155	27.38
feel_o_4	1118	26.50
feel_o_5	1130	26.78
feel_o_6	1146	27.16
feel_o_7	1164	27.59
feel_o_8	1158	27.45
feel_o_9	1174	27.83
perc_o_1	1143	27.09
perc_o_2	1149	27.23
perc_o_3	1167	27.66
perc_o_4	1167	27.66
perc_o_5	1186	28.11
perc_o_6	1182	28.02
perc_o_7	1192	28.25

Continued on next page

Table C.4 Missings Prior to EM – *Continued*

	n	%
perc_o_8	1192	28.25
underst_o_1	1415	33.54
underst_o_2	1402	33.23
underst_o_3	1409	33.40
underst_o_4	1413	33.49
underst_o_5	1415	33.54
underst_o_6	1404	33.28
underst_o_7	1415	33.54
underst_o_8	1455	34.49
cogabs_1	1380	32.71
cogabs_2	1382	32.76
cogabs_3	1390	32.95
cogabs_4	1390	32.95
cogabs_5	1393	33.02
cogabs_6	1412	33.47
cogabs_7	1403	33.25
cogabs_8	1414	33.52
feel_man_1	1509	35.77
feel_man_2	1510	35.79
feel_man_3	1505	35.67
feel_man_4	1509	35.77
feel_man_5	1512	35.84
feel_man_6	1540	36.50
feel_man_7	1580	37.45
perc_s_1	1608	38.11
underst_s_2	1562	37.02
underst_s_3	1569	37.19
perc_s_2	1623	38.47
perc_s_3	1633	38.71
perc_s_4	1662	39.39
underst_s_1	1658	39.30

D

RESULTS OF ANALYSES

D.1 EXPLORATORY FACTOR ANALYSES

D.1.1 *Suitability Tests and Factor Structures*

In this section, we present the results of our EFAs ($N = 4{,}219$) which we performed on the basis of the initial item set before eliminating problematic indicators. The number of factors to retain were determined with the aid of PA and the MAP test. First, the KMO measure and results of the Bartlett test are displayed (i. e., measures which indicate the extent to which our data is suitable for factor analysis), for each construct separately and grouped according to their initial scale formations and attribution of items with regard to their hypothesized constructs. Second, we display obtained MSA values for all items; furthermore, for items belonging to constructs found to be multidimensional, we additionally display individual factor loadings after ML extraction with promax rotation (in form of the pattern matrices)–ordered by factors, where boldface indicates highest factor loadings. This is then followed by the factor correlations for all factors of the construct. Note that the results of Tables D.1-D.3 were obtained after elimination of all reverse coded items of the IR emotional capabilities construct ($underst\_o\_6$, $underst\_o\_7$, $feel\_o\_6$, $feel\_o\_3$, $feel\_man\_2$, $feel\_man\_1$, see Section 4.4).

Primary Constructs

Table D.1: KMO and Bartlett's Test for IR Emotional Capabilities

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.942
Bartlett's Test of Sphericity:	Approx. χ^2	81732.641
	df	528
	Sig.	.000

Table D.2: Eigenvalues and Percentages of Variance for Factors of IR Emotional Capabilities

	EV	% of Var.
Factor 1	11.297	34.232
Factor 2	3.371	10.216
Factor 3	2.449	7.421
Factor 4	1.638	4.963
Factor 5	1.305	3.956
% of variance explained by the 5 factors		60.787

Table D.3: MSA and Factor Loadings for IR Emotional Capabilities

Item	MSA	Factor loadings 1	2	3	4	5
perc_o_6	.953	**1.017**				-.164
perc_o_5	.958	**.948**				
perc_o_7	.948	**.942**	.100			-.203
perc_o_8	.975	**.846**				
perc_o_4	.965	**.770**				
perc_o_2	.940	**.618**				.144
perc_o_3	.985	**.553**				
perc_o_1	.941	**.503**				.197
underst_o_4	.962	**.441**				.384
underst_o_8	.979	**.350**	.170			.171
feel_o_8	.874		**.932**			
feel_o_9	.882		**.889**			
feel_o_7	.937		**.857**			
feel_o_4	.938		**.676**			
feel_o_2	.946		**.654**			
feel_o_5	.960	.162	**.579**			
feel_o_1	.959	.116	**.210**	.121		.201
feel_man_5	.861			**.920**		
feel_man_4	.907			**.741**		
feel_man_3	.911			**.740**		
feel_man_6	.927			**.684**		
feel_man_7	.961			**.415**	.153	
underst_s_2	.893				**.978**	
underst_s_3	.918				**.864**	
perc_s_1	.950				**.716**	
perc_s_2	.909			.166	**.343**	
underst_s_1	.970	.176			**.332**	
perc_s_3	.915	.179	.110	.142	**.244**	
perc_s_4	.907		.181		**.187**	
underst_o_2	.933					**.866**
underst_o_1	.928					**.864**
underst_o_3	.967	.396				**.488**
underst_o_5	.970	.351				**.385**

Extraction method: Maximum likelihood.
Rotation method: Promax with Kaiser normalization.
Rotation converged in 6 iterations.

Table D.4: Factor Correlations for IR Emotional Capabilities

	1	2	3	4	5
Factor 1	-				
Factor 2	.460	-			
Factor 3	.360	.145	-		
Factor 4	.557	.303	.505	-	
Factor 5	.684	.417	.398	.606	-

Note. MSA and the factor loadings are presented on p. 238.

Table D.5: KMO and Bartlett's Test for SE Cognitive Ability

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.856
Bartlett's Test of Sphericity:	Approx. χ^2	14763.942
	df	21
	Sig.	.000

Table D.6: Eigenvalues and Percentages of Variance for Factors of SE Cognitive Ability

	EV	% of Var.
Factor 1	3.953	56.477
Factor 2	1.139	16.277
% of variance explained by the 2 factors		72.754

Table D.7: MSA, Factor Loadings, and Factor Correlations for SE Cognitive Ability

		Factor	
Item	**MSA**	1	2
clever_6	.832	**.897**	-.103
clever_4	.854	**.820**	
clever_7	.894	**.674**	
clever_5	.910	**.554**	.298
clever_2	.804		**.939**
clever_3	.833		**.774**
clever_1	.873	.263	**.606**
		Factor correlations	
	Factor 1	-	
	Factor 2	.631	-

Extraction method: Maximum likelihood.
Rotation method: Promax with Kaiser normalization.
Rotation converged in 3 iterations.

Table D.8: KMO and Bartlett's Test for CIRME

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.783
Bartlett's Test of Sphericity:	Approx. χ^2	27877.127
	df	66
	Sig.	.000

Table D.9: Eigenvalues and Percentages of Variance for Factors of CIRME

	EV	% of Var.
Factor 1	4.194	34.952
Factor 2	1.991	16.596
Factor 3	1.789	14.908
Factor 4	1.358	11.321
% of variance explained by the 4 factors		77.777

Table D.10: MSA, Factor Loadings, and Factor Correlations for CIRME

		Factor loadings			
Item	**MSA**	1	2	3	4
internwo10_14_2	.742	**.968**			
internwo10_14_1	.776	**.918**			
internwo10_14_3	.879	**.822**			
wothers10_14_2	.785		**.885**		
wothers10_14_3	.805		**.837**		
wothers10_14_1	.861		**.740**		
allone10_14_2	.762			**.953**	
allone10_14_1	.746			**.827**	
allone10_14_3	.877	.101		**.574**	
watch10_14_1	.659				**.817**
watch10_14_2	.667				**.794**
watch10_14_3	.779				**.566**
		Factor correlations			
	Factor 1	-			
	Factor 2	.308	-		
	Factor 3	.368	.444	-	
	Factor 4	.109	.166	.192	-

Extraction method: Maximum likelihood.
Rotation method: Promax with Kaiser normalization.
Rotation converged in 5 iterations.

Table D.11: KMO and Bartlett's Test for Cognitive Absorption

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.864
Bartlett's Test of Sphericity:	Approx. χ^2	16106.076
	df	28
	Sig.	.000

Table D.12: Eigenvalues and Percentages of Variance for Factors of Cognitive Absorption

	EV	% of Var.
Factor 1	4.194	34.952
Factor 2	1.991	16.596
% of variance explained by the 2 factors		51.548

Table D.13: MSA, Factor Loadings, and Factor Correlations for Cognitive Absorption

		Factor	
Item	**MSA**	1	2
cogabs_1	.854	**.900**	-.105
cogabs_2	.867	**.887**	
cogabs_8	.869	**.798**	
cogabs_3	.893	**.671**	
cogabs_5	.845		**.810**
cogabs_4	.868	.203	**.642**
cogabs_7	.869		**.610**
cogabs_6	.416	-.103	**.117**
		Factor correlations	
	Factor 1	-	
	Factor 2	.667	-

Extraction method: Maximum likelihood.
Rotation method: Promax with Kaiser normalization.
Rotation converged in 3 iterations.

As can be seen from Table D.13 on p. 241, the *cogabs_6* item was just below the MSA threshold of .50; the item was reversely coded, a characteristic which may have been overlooked by participants. It was also found not to load significantly on any of the two factors extracted for this scale, and to lack reliability.

Secondary Constructs

Table D.14: KMO and Bartlett's Test for Self-Motivational Traits

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.864
Bartlett's Test of Sphericity:	Approx. χ^2	13022.148
	df	21
	Sig.	.000

Table D.15: Eigenvalues and Percentages of Variance for Factors of Self-Motivational Traits

	EV	% of Var.
Factor 1	3.845	54.926
Factor 2	1.064	15.205
% of variance explained by the 2 factors		70.131

Table D.16: MSA, Factor Loadings, and Factor Correlations for Self-Motivational Traits

		Factor	
Item	**MSA**	1	2
motivation_4	.835	**.872**	
motivation_3	.841	**.847**	
motivation_2	.910	**.675**	
motivation_1	.911	**.597**	.192
pers_stand_5	.835		**.837**
pers_stand_4	.848		**.756**
pers_stand_6	.894		**.604**
		Factor correlations	
	Factor 1	-	
	Factor 2	.638	-

Extraction method: Maximum likelihood.
Rotation method: Promax with Kaiser normalization.
Rotation converged in 3 iterations.

Table D.17: KMO and Bartlett's Test for Competitiveness

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.929
Bartlett's Test of Sphericity:	Approx. χ^2	45676.013
	df	120
	Sig.	.000

Table D.18: Eigenvalues and Percentages of Variance for Factors of Competitiveness

	EV	% of Var.
Factor 1	7.312	45.698
Factor 2	2.691	16.819
Factor 3	1.259	7.867
% of variance explained by the 3 factors		70.384

Table D.19: MSA, Factor Loadings, and Factor Correlations for Competitiveness

		Factor loadings		
Item	**MSA**	1	2	3
perf_orient_5	.943	**.846**		
perf_orient_7	.939	**.820**		
perf_orient_3	.945	**.814**		
perf_orient_2	.937	**.777**		
perf_orient_6	.938	**.769**		
perf_orient_1	.947	**.768**		
perf_orient_4	.944	**.767**		
perf_orient_8	.947	**.629**		
perfect_2	.902		**.936**	
perfect_3	.911		**.915**	
perfect_5	.936		**.852**	
perfect_1	.931		**.831**	
perfect_4	.951		**.780**	
pers_stand_3	.973	.105	**.374**	.301
pers_stand_2	.836			**.877**
pers_stand_1	.827			**.771**
		Factor correlations		
	Factor 1	-		
	Factor 2	.459	-	
	Factor 3	.412	.514	-

Extraction method: Maximum likelihood.
Rotation method: Promax with Kaiser normalization.
Rotation converged in 3 iterations.

Table D.20: KMO and Bartlett's Test for IR Enjoyment

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.732
Bartlett's Test of Sphericity:	Approx. χ^2	11401.701
	df	6
	Sig.	.000

Table D.21: MSA for IR Enjoyment

Item	MSA
fun_1	.678
fun_2	.678
fun_3	.830
fun_4	.809

Table D.22: KMO and Bartlett's Test for IR Mediation

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.861
Bartlett's Test of Sphericity:	Approx. χ^2	16579.531
	df	10
	Sig.	.000

Table D.23: MSA for IR Mediation

Item	MSA
mediation_1	.817
mediation_2	.815
mediation_3	.892
mediation_4	.912
mediation_5	.915

Table D.24: KMO and Bartlett's Test for Parental Control

Kaiser-Meyer-Olkin Measure of Sampling Adequacy:		.768
Bartlett's Test of Sphericity:	Approx. χ^2	7276.280
	df	10
	Sig.	0.000

Table D.25: MSA for Parental Control

Item	MSA
rules_parents_1	.630
rules_parents_2	.754
rules_parents_3	.792
rules_parents_4	.799
rules_parents_5	.754

D.1.2 *Preliminary Reliability Analyses*

Primary Constructs

Table D.26: Cronbach's Alpha and Inter-Item Statistics for IR Emotional Capabilities

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	perc_o_6	.921 (.922)	.541	.808	.716	.907
	perc_o_5			.818	.713	.907
	perc_o_7			.743	.638	.911
	perc_o_8			.775	.627	.909
	perc_o_4			.717	.577	.913
	perc_o_2			.691	.600	.914
	perc_o_3			.619	.389	.919
	perc_o_1			.663	.572	.915
	underst_o_4			.676	.481	.915
	underst_o_8			.532	.344	.922
2	feel_o_8	.885 (.882)	.517	.755	.707	.858
	feel_o_9			.715	.660	.863
	feel_o_7			.759	.619	.857
	feel_o_4			.700	.525	.865
	feel_o_2			.709	.519	.864
	feel_o_5			.685	.498	.867

Continued on next page

Table D.26 Alpha & Inter-Item Statistics for IR Emotional Capabilities – *Continued*

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
	feel_o_1			.396	.236	.898
3	feel_man_5			.738	.570	.760
	feel_man_4			.659	.463	.783
	feel_man_3	.827 (.829)	.491	.628	.435	.792
	feel_man_6			.620	.391	.794
	feel_man_7			.485	.244	.833
4	underst_s_2			.629	.636	.764
	underst_s_3			.610	.590	.768
	perc_s_1			.598	.477	.769
	perc_s_2	.805 (.811)	.380	.579	.398	.772
	underst_s_1			.534	.293	.780
	perc_s_3			.529	.369	.781
	perc_s_4			.338	.165	.820
5	underst_o_2			.763	.637	.801
	underst_o_1	.862 (.862)	.610	.720	.605	.818
	underst_o_3			.718	.526	.819
	underst_o_5			.636	.439	.853

Note. Items were assigned to factors according to ML extraction with promax rotation. *underst_o_6*, *underst_o_7*, *feel_o_3*, *feel_o_6*, *feel_man_1*, and *feel_man_2* were excluded.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.27: Cronbach's Alpha and Inter-Item Statistics for SE Cognitive Ability

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	clever_1			.650	.457	.785
	clever_2	.823 (.823)	.608	.756	.572	.675
	clever_3			.632	.429	.801
2	clever_4			.747	.598	.763
	clever_5	.838 (.842)	.572	.645	.462	.807
	clever_6			.745	.573	.761
	clever_7			.567	.338	.847

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.28: Cronbach's Alpha and Inter-Item Statistics for CIRME

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	internwo10_14_1			.865	.779	.891
	internwo10_14_2	.930 (.930)	.815	.894	.811	.867
	internwo10_14_3			.810	.662	.934
2	wothers10_14_1			.712	.510	.835
	wothers10_14_2	.864 (.864)	.679	.772	.597	.780
	wothers10_14_3			.742	.559	.809
3	allone10_14_1			.708	.586	.742
	allone10_14_2	.828 (.832)	.622	.774	.635	.676
	allone10_14_3			.589	.360	.865
4	watch10_14_1			.656	.459	.612
	watch10_14_2	.763 (.763)	.518	.637	.446	.633
	watch10_14_3			.498	.249	.786

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.29: Cronbach's Alpha and Inter-Item Statistics for Cognitive Absorption

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if item del.[d]
1	cogabs_1	.871 (.873)	.633	.749	.605	.827
	cogabs_2			.782	.613	.811
	cogabs_3			.649	.457	.868
	cogabs_8			.730	.582	.833
2	cogabs_4	.634 (.657)	.324	.548	.500	.465
	cogabs_5			.611	.555	.431
	cogabs_6			.058	.019	.815
	cogabs_7			.564	.378	.452

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Secondary Constructs

Table D.30: Cronbach's Alpha and Inter-Item Statistics for Self-Motivational Traits

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	motivation_1	.851 (.853)	.592	.661	.446	.823
	motivation_2			.615	.380	.846
	motivation_3			.746	.589	.787
	motivation_4			.751	.599	.785
2	pers_stand_4	.782 (.787)	.552	.628	.415	.699
	pers_stand_5			.665	.452	.660
	pers_stand_6			.580	.339	.761

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.31: Cronbach's Alpha and Inter-Item Statistics for Competitiveness

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	perf_orient_1			.751	.610	.911
	perf_orient_2			.742	.602	.912
	perf_orient_3			.783	.649	.909
	perf_orient_4	.922 (.923)	.599	.691	.521	.916
	perf_orient_5			.795	.645	.908
	perf_orient_6			.742	.595	.912
	perf_orient_7			.770	.612	.910
	perf_orient_8			.643	.470	.920
2	perfect_1			.773	.621	.923
	perfect_2			.851	.749	.908
	perfect_3	.931 (931)	.729	.850	.744	.909
	perfect_4			.800	.647	.918
	perfect_5			.812	.667	.916
3	pers_stand_1			.620	.450	.621
	pers_stand_2	.751 (.756)	.507	.655	.473	.581
	pers_stand_3			.474	.228	.794

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.32: Cronbach's Alpha and Inter-Item Statistics for IR Enjoyment

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	fun_1	.870 (.875)	.637	.762	.797	.820
	fun_2			.802	.813	.805
	fun_3			.733	.545	.829
	fun_4			.619	.431	.880

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.33: Cronbach's Alpha and Inter-Item Statistics for IR Mediation

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	mediation_1			.833	.803	.874
	mediation_2			.845	.810	.872
	mediation_3	.908 (.910)	.670	.844	.742	.873
	mediation_4			.740	.554	.894
	mediation_5			.601	.384	.924

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

Table D.34: Cronbach's Alpha and Inter-Item Statistics for Parental Control

Factor	Item	Cronbach's α (stand.)	MIC[a]	Corr. ITC[b]	SMC[c]	α if it. del.[d]
1	rules_parents_1			.207	.084	.840
	rules_parents_2			.678	.522	.691
	rules_parents_3	.778 (.776)	.410	.673	.509	.696
	rules_parents_4			.610	.426	.716
	rules_parents_5			.630	.523	.709

Note. Items were assigned to factors according to ML extraction with promax rotation.
[a] Mean inter-item correlation.
[b] Corrected item-total correlation.
[c] Squared multiple correlation.
[d] Cronbach's α if item deleted.

D.1.3 *Simultaneous Exploratory Factor Analysis*

The following three tables show the results of three ML-based EFAs which included all primary construct variables and used promax rotation with a predefined 12-, 11-, and 10-factor structure. All reversely coded items of IR emotional capabilities, that is, *underst_o_6*, *underst_o_7*, *feel_o_3*, *feel_o_6*, *feel_man_1*, and *feel_man_2*, as well as problematic items in terms or reliability and loadings of the same construct, *underst_o_4*, *underst_o_8*, *feel_o_1*, *feel_man_7*, *perc_s_2*, *underst_s_1*, *perc_s_3*, *perc_s_4*, and *underst_o_5*, and finally *cogabs_6* of cognitive absorption and *clever_5* of SE cognitive ability were excluded; see Section 4.4 for details.

Table D.35: Factor Loadings of Primary Construct Indicators with 12-Factor Structure

	Factor											
	1	2	3	4	5	6	7	8	9	10	11	12
perc_o_6	**.956**											-.101
perc_o_5	**.920**											
perc_o_7	**.858**					.154				-.164		-.108
perc_o_8	**.810**											
perc_o_4	**.781**									.109		
perc_o_2	**.659**					-.122				.146		
perc_o_3	**.556**											.102
perc_o_1	**.555**					-.157				.226		.116

Continued on next page

Table D.35 Simultaneous EFA - 12-Factor Structure – *Continued*

	Factor											
	1	2	3	4	5	6	7	8	9	10	11	12
cogabs_2		**.846**										
cogabs_8		**.825**										
cogabs_1		**.768**				-.119				.217		
cogabs_3		**.759**										
cogabs_4		**.688**				.114				-.117		
cogabs_5		**.647**										
cogabs_7		**.524**				.144				-.185		
feel_o_8			**.910**									
feel_o_9			**.855**							-.123		
feel_o_7			**.848**									
feel_o_4			**.700**							.144		
feel_o_2			**.635**									
feel_o_5	.174		**.579**									
internwo10_14_2				**.967**								
internwo10_14_1				**.924**								
internwo10_14_3				**.818**								
feel_man_5					**.899**							
feel_man_3					**.724**							
feel_man_4					**.714**							
feel_man_6					**.647**							

Continued on next page

Table D.35 Simultaneous EFA - 12-Factor Structure – *Continued*

	Factor											
	1	2	3	4	5	6	7	8	9	10	11	12
clever_6						**.801**				.138		
clever_4						**.750**				.267		
clever_7						**.609**						
wothers10_14_2							**.881**					
wothers10_14_3							**.845**					
wothers10_14_1							**.736**					
allone10_14_2								**.946**				
allone10_14_1								**.854**		.106		
allone10_14_3								**.566**		-.160		
underst_s_2									**.964**			
underst_s_3									**.791**			
perc_s_1									**.669**			
clever_2						.179				**.749**		
clever_1						.371				**.667**		
clever_3										**.619**		
watch10_14_1											**.806**	
watch10_14_2											**.803**	
watch10_14_3						.169				-.161	**.568**	

Continued on next page

Table D.35 Simultaneous EFA - 12-Factor Structure – *Continued*

	Factor											
	1	2	3	4	5	6	7	8	9	10	11	12
underst_o_2	.115											**.842**
underst_o_1												**.827**
underst_o_3	.418											**.422**

Extraction Method: Maximum Likelihood. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 7 iterations.

Table D.36: Factor Loadings of Primary Construct Indicators with 11-Factor Structure

	Factor										
	1	2	3	4	5	6	7	8	9	10	11
perc_o_6	**.941**									-.152	
perc_o_5	**.915**										
perc_o_8	**.847**										
perc_o_7	**.840**					.177				-.235	
perc_o_4	**.785**										
perc_o_2	**.730**					-.122				.164	
perc_o_1	**.631**					-.154				.251	
perc_o_3	**.620**										

Continued on next page

Table D.36 Simultaneous EFA - 11-Factor Structure – *Continued*

	Factor										
	1	2	3	4	5	6	7	8	9	10	11
underst_o_3	**.600**										
underst_o_2	**.451**								.134	.197	
underst_o_1	**.387**		.128						.124	.241	
cogabs_2		**.846**									
cogabs_8		**.825**									
cogabs_1		**.766**				-.106				.226	
cogabs_3		**.758**									
cogabs_4		**.673**									
cogabs_5		**.631**							.119		
cogabs_7		**.512**				.118				-.162	

Continued on next page

Table D.36 Simultaneous EFA - 11-Factor Structure – *Continued*

	Factor										
	1	2	3	4	5	6	7	8	9	10	11
feel_o_8			**.915**								
feel_o_9			**.864**							-.146	
feel_o_7			**.850**								
feel_o_4			**.695**							.134	
feel_o_2			**.647**								
feel_o_5	.183		**.578**								
internwo10_14_2				**.967**							
internwo10_14_1				**.924**							
internwo10_14_3				**.817**							
feel_man_5					**.897**						
feel_man_3					**.721**						
feel_man_4					**.711**						
feel_man_6					**.645**						
clever_6						**.778**				.189	
clever_4						**.733**				.327	
clever_7						**.602**				.112	
wothers10_14_2							**.887**				
wothers10_14_3							**.853**				
wothers10_14_1							**.741**				

Continued on next page

Table D.36 Simultaneous EFA - 11-Factor Structure – *Continued*

	Factor										
	1	2	3	4	5	6	7	8	9	10	11
allone10_14_2								**.945**			
allone10_14_1								**.850**			
allone10_14_3								**.565**		-.148	
underst_s_2									**.962**		
underst_s_3									**.804**		
perc_s_1									**.705**		
clever_2						.220				**.754**	
clever_1						.398				**.695**	
clever_3						.127				**.620**	
watch10_14_1											**.809**
watch10_14_2											**.807**
watch10_14_3						.167				-.182	**.575**

Extraction Method: Maximum Likelihood. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 7 iterations.

Table D.37: Factor Loadings of Primary Construct Indicators with 10-Factor Structure

	Factor									
	1	2	3	4	5	6	7	8	9	10
perc_o_6	**.951**							-.106		
perc_o_5	**.931**									
perc_o_8	**.857**									
perc_o_7	**.849**							-.160		
perc_o_4	**.792**									
perc_o_2	**.720**									
perc_o_3	**.624**									
perc_o_1	**.615**									
underst_o_3	**.600**							.115		
underst_o_2	**.440**							.221		
underst_o_1	**.375**							.224		
cogabs_2		**.886**								
cogabs_8		**.849**								
cogabs_1		**.834**								
cogabs_3		**.776**								
cogabs_4		**.641**								
cogabs_5		**.603**								
cogabs_7		**.456**								

Continued on next page

Table D.37 Simultaneous EFA - 10-Factor Structure – *Continued*

	Factor									
	1	2	3	4	5	6	7	8	9	10
feel_o_8			**.918**							
feel_o_9			**.875**							
feel_o_7			**.839**							
feel_o_4			**.662**					.101		
feel_o_2			**.630**							
feel_o_5	.183		**.565**							
clever_4				**.885**						
clever_6				**.841**						
clever_1				**.679**						
clever_7				**.639**						
clever_2		.110		**.522**				.118		
clever_3		.142		**.385**				.183		
internwo10_14_2					**.964**					
internwo10_14_1					**.915**					
internwo10_14_3					**.816**					
feel_man_5						**.918**				
feel_man_4						**.735**				
feel_man_3						**.721**				
feel_man_6						**.649**				

Continued on next page

Table D.37 Simultaneous EFA - 10-Factor Structure – *Continued*

	Factor									
	1	2	3	4	5	6	7	8	9	10
wothers10_14_2							**.909**			
wothers10_14_3							**.854**			
wothers10_14_1							**.759**			
underst_s_2								**.970**		
underst_s_3								**.845**		
perc_s_1								**.724**		
allone10_14_2									**.915**	
allone10_14_1									**.835**	
allone10_14_3					.100				**.583**	.106
watch10_14_1										**.803**
watch10_14_2										**.790**
watch10_14_3									.101	**.580**

Extraction Method: Maximum Likelihood. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 7 iterations.

D.2 Confirmatory Factor Analyses

D.2.1 *Model Fit Indexes and Minimal Sample Sizes*

We compare (meaningful) candidate measurement models in terms of fit indexes for each construct below and further contrast them with their respective null model. The legend below indicates the recommended threshold values in parentheses. All problematic indicators were excluded from this analysis (see Section 4.4 and the previous section). On the basis of a sample size of 4,048 (cf. the minimal sample size column for each construct), all χ^2 statistics were significant ($p < .000$). The probability of getting critical ratios as large as indicated in absolute value was less than .001 (i. e., $p < .001$).

Legend:

χ^2	χ^2 statistics of the respective model
RMSEA	Root mean square error of approximation ($\leq .06$)
NNFI	Nonnormed fit index ($\geq .95$)
CFI	Comparative fit index ($\geq .95$)
SRMR	Standardized root mean square residual ($\leq .08$)
Min. N	Minimal sample size according to MacCallum et al. (1996) ($\alpha = .01$, power $= .99$; close fit, i. e., $\epsilon_0 = .05$, $\epsilon_a = .08$)

Primary Constructs

Table D.38: Model Fit Indexes and Minimal Sample Size for Models of IR Emotional Capabilites

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	43,445.077	300					
Four factor	5,692.041	246	.074	.846	.874	.0645	169
Second-order four factor	5,832.274	248	.075	.843	.871	.0712	169
Five factor	3,911.419	242	.061	.895	.915	.0586	172
Second-order five factor	4,055.834	247	.062	.893	.912	.0660	169

Note. Included perc_o_1, perc_o_3, perc_o_2, perc_o_4, perc_o_8, perc_o_7, perc_o_5, perc_o_6, feel_o_5, feel_o_2, feel_o_4, feel_o_7, feel_o_9, feel_o_8, feel_man_6, feel_man_4, feel_man_3, feel_man_5, perc_s_1, underst_s_3, underst_s_2, underst_o_3, underst_o_1, and underst_o_2.

Table D.39: Model Fit Indexes and Minimal Sample Size for Models of SE Cognitive Ability

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	8,959.510	21					
One factor	1,977.563	9	.232	.486	.780	.1015	2800
Two factor	476.134	8	.120	.863	.948	.0517	3150

Note. Included clever_7, clever_6, clever_4, clever_3, clever_2, and clever_1.

Table D.40: Model Fit Indexes and Minimal Sample Size for Models of CIRME

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	19,566.967	78					
Second-order four factor	467.566	50	.045	.967	.979	.0438	575
Four factor	465.140	48	.046	.965	.979	.0436	607

Note. Included *watch*10_14_3, *watch*10_14_2, *watch*10_14_1, *allone*10_14_3, *allone*10_14_2, *allone*10_14_1, *wothers*10_14_3, *wothers*10_14_2, *wothers*10_14_1, *internwo*10_14_3, *internwo*10_14_2, and *internwo*10_14_1.

Table D.41: Model Fit Indexes and Minimal Sample Size for Models of Cognitive Absorption

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	10,662.158	28					
One factor	1,650.026	14	.170	.692	.846	.0779	1850
Two factor	1,625.142	13	.175	.673	.848	.0779	1950

Note. Included *cogabs*_8, *cogabs*_4, *cogabs*_3, *cogabs*_2, *cogabs*_1, *cogabs*_5, and *cogabs*_7.

Secondary Constructs

Table D.42: Model Fit Indexes and Minimal Sample Size for Models of Self-Motivational Traits

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	8448,516	28					
One factor	1239,232	14	.147	.709	.854	.0841	1850
Two factor	111.735	13	.043	.975	.988	.0263	1950

Note. Included *pers_stand*_6, *pers_stand*_5, *pers_stand*_4, *motivation*_4, *motivation*_3, *motivation*_2, and *motivation*_1.

Table D.43: Model Fit Indexes and Minimal Sample Size for Models of Competitiveness

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	28,628.794	91					
Two factor	1,414.228	64	.072	.933	.953	.0387	475

Note. Included *perfect*_5, *perfect*_4, *perfect*_3, *perfect*_2, *perfect*_1, *perf_orient*_8, *perf_orient*_7, *perf_orient*_6, *perf_orient*_5, *perf_orient*_4, *perf_orient*_3, *perf_orient*_2, and *perf_orient*_1.

Table D.44: Model Fit Indexes and Minimal Sample Size for Models of IR Enjoyment

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	7,433.586	10					
One factor	561.048	2	.263	.623	.925	.0824	12,100[a]

Note. Included *fun*_4, *fun*_3, *fun*_2, and *fun*_1.
[a] Power with actual sample size of 4,048: .64.

Table D.45: Model Fit Indexes and Minimal Sample Size for Models of IR Mediation

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	8,557.159	15					
One factor	319.325	5	.125	.890	.963	.0459	4,950[a]

Note. Included *mediation*_5, *mediation*_4, *mediation*_3, *mediation*_2, and *mediation*_1.
[a] Power with actual sample size of 4,048: .97.

Table D.46: Model Fit Indexes and Minimal Sample Size for Models of Parental Control

Model	χ^2	*df*	RMSEA	NNFI	CFI	SRMR	Min. N
Null	6,179.140	15					
One factor	444.369	5	.147	.786	.929	.047	4950[a]

Note. Included *rules_parents_5*, *rules_parents_4*, *rules_parents_3*, *rules_parents_2*, and *rules_parents_1*.
[a] Power with actual sample size of 4,048: .97.

D.2.2 *Loadings and Validity / Reliability Indexes*

Only results for the primary constructs are presented. Critical ratios of all estimates (i. e., hypothesized relationships between indicators and factors as well as of variances and intercepts, $N = 4{,}048$) were significant at the .001 level ($p < .001$, tested via two-tailed z-tests, cf. MacKenzie et al., 2011). The legend below indicates the recommended threshold values for the respective index in parentheses. The corresponding formulas have been introduced in Section 4.6.2.

Legend:

λ_i	estimated factor loading of an indicator with its hypothesized factor i ($\geq .50$)
ρ_{ind}	squared multiple correlation of an indicator ($\geq .40$)
ρ_{comp}	composite reliability of a factor or construct ($\geq .60$)
AVE	average variance extracted of a factor or construct ($\geq .50$)
Φ_{ij}^2	squared correlation between factors ξ_i and ξ_j

Primary Constructs

Table D.47: Factor Loadings and Item/Factor Reliability and Validity for IR Emotional Capabilities

		Factor loadings					Reliability		Validity
Factor	Item	λ_1	λ_2	λ_3	λ_4	λ_5	ρ_{ind}	ρ_{comp}	AVE
F1	perc_o_5	.874					.764		
	perc_o_6	.866					.750		
	perc_o_8	.820					.673		
	perc_o_7	.783					.613	.921	.597
	perc_o_4	.772					.596		
	perc_o_2	.711					.506		
	perc_o_1	.687					.473		
	perc_o_3	.632					.400		
F2	feel_o_8		.864				.747		
	feel_o_7		.831				.691		
	feel_o_9		.825				.680	.867	.602
	feel_o_4		.721				.520		
	feel_o_2		.702				.492		
	feel_o_5		.694				.481		

Continued on next page

Table D.47 Loadings and Reliability - IR Emotional Capabilities – *Continued*

		Factor loadings					Reliability		Validity
Factor	Item	λ_1	λ_2	λ_3	λ_4	λ_5	ρ_{ind}	ρ_{comp}	AVE
F3	feel_man_5			.865			.749	.768	.580
	feel_man_4			.757			.573		
	feel_man_3			.738			.545		
	feel_man_6			.674			.454		
F4	underst_s_2				.896		.802	.720	.680
	underst_s_3				.838		.701		
	perc_s_1				.733		.538		
F5	underst_o_2					.883	.780	.726	.682
	underst_o_1					.844	.712		
	underst_o_3					.744	.554		

Note. Φ^2_{12}: .161; Φ^2_{13}: .075; Φ^2_{14}: .234; Φ^2_{15}: .475; Φ^2_{23}: .004; Φ^2_{24}: .049; Φ^2_{25}: .140; Φ^2_{34}: .158; Φ^2_{35}: .110; Φ^2_{45}: .297.

Table D.48: Factor Loadings and Item/Factor Reliability and Validity for SE Cognitive Ability

		Factor loadings		Reliability		Validity
Factor	Item	λ_1	λ_2	ρ_{ind}	ρ_{comp}	AVE
1 Quotidian feedback	clever_2	.843		.710		
	clever_1	.794		.631	.820	.605
	clever_3	.688		.473		
2 Institutional feedback	clever_4		.857	.734		
	clever_6		.835	.697	.818	.604
	clever_7		.617	.380		

Note. Φ^2_{12}: .417.

Table D.49: Factor Loadings and Item/Factor Reliability and Validity for CIRME

		Factor loadings				Reliability		Validity
Factor	Item	λ_1	λ_2	λ_3	λ_4	ρ_{ind}	ρ_{comp}	AVE
1 Watch[a]	watch10_14_1	.813				.660		
	watch10_14_2	.797				.635	.769	.532
	watch10_14_3	.549				.301		
2 Alone	allone10_14_2		.923			.851		
	allone10_14_1		.823			.677	.842	.645
	allone10_14_3		.638			.407		
3 With others	wothers10_14_2			.869		.754		
	wothers10_14_3			.817		.667	.880	.725
	wothers10_14_1			.776		.603		
4 Over network	internwo10_14_2				.956	.915		
	internwo10_14_1				.916	.838	.930	.816
	internwo10_14_3				.834	.695		

Note. Φ^2_{12}: .026; Φ^2_{13}: .020; Φ^2_{14}: .005; Φ^2_{23}: .178; Φ^2_{24}: .101; Φ^2_{34}: .072.

[a] ρ_{comp} changed to .786 and AVE to .648 after excluding *watch*10_14_3 from the calculations.

Table D.50: Factor Loadings and Item/Factor Reliability and Validity for Cognitive Absorption

		Factor loadings	Reliability		Validity
Factor	Item	λ_1	ρ_{ind}	ρ_{comp}	AVE
1	cogabs_8	.822	.676		
	cogabs_2	.808	.652		
	cogabs_1	.796	.633		
	cogabs_3	.709	.503	.888	.534
	cogabs_5	.700	.489		
	cogabs_4	.690	.476		
	cogabs_7	.559	.312		

[a] AVE changed to .572 after excluding *cogabs_7* from the calculations.

D.3 STRUCTURAL EQUATION MODELING

D.3.1 *Correlations, Means, and Standard Deviations*

Table D.51: Intercorrelations, Means, and Standard Deviations of Variables Included in the Analysis I - ML estimation with original performance data

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Understand others' IR emotions																								
1. underst_o_3	–																							
2. underst_o_1	.62	–																						
3. underst_o_2	.64	.72	–																					
Perceive / understand own IR emotions																								
4. perc_s_1	.27	.31	.32	–																				
5. underst_s_3	.31	.36	.37	.60	–																			
6. underst_s_2	.34	.38	.40	.65	.74	–																		
Regulate own IR emotions																								
7. feel_man_6	.16	.18	.18	.19	.22	.23	–																	
8. feel_man_3	.17	.19	.20	.21	.24	.25	.50	–																
9. feel_man_4	.17	.20	.20	.21	.24	.26	.51	.56	–															
10. feel_man_5	.20	.23	.23	.24	.28	.30	.58	.64	.65	–														
Act on others' IR emotions																								
11. feel_o_5	.20	.22	.23	.10	.12	.12	.06	.07	.07	.08	–													
12. feel_o_2	.20	.23	.23	.10	.12	.13	.06	.07	.07	.08	.48	–												
13. feel_o_4	.21	.23	.24	.11	.12	.13	.06	.07	.07	.08	.50	.50	–											
14. feel_o_9	.23	.27	.28	.12	.14	.15	.07	.08	.08	.09	.57	.58	.59	–										
15. feel_o_7	.24	.27	.28	.12	.14	.15	.07	.08	.08	.09	.57	.58	.60	.68	–									
16. feel_o_8	.25	.28	.29	.13	.15	.16	.08	.08	.09	.10	.60	.60	.62	.71	.72	–								

Continued on next page

Table D.51 Correlations, Means, and Standard Deviations I - ML (untransformed data) – *Continued*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Perceive others' IR emotions																								
17. perc_o_3	.31	.35	.36	.19	.22	.23	.10	.11	.11	.13	.11	.11	.12	.13	.13	.14	–							
18. perc_o_1	.34	.38	.39	.21	.24	.25	.11	.12	.12	.14	.12	.12	.13	.14	.15	.15	.43	–						
19. perc_o_2	.35	.39	.41	.21	.25	.26	.11	.12	.12	.14	.13	.13	.13	.15	.15	.16	.45	.49	–					
20. perc_o_4	.38	.43	.44	.23	.27	.29	.12	.13	.13	.15	.14	.14	.14	.16	.16	.17	.49	.53	.55	–				
21. perc_o_7	.38	.43	.45	.24	.27	.29	.12	.13	.14	.16	.14	.14	.14	.16	.17	.17	.49	.54	.56	.60	–			
22. perc_o_8	.40	.45	.47	.25	.28	.30	.13	.14	.14	.16	.15	.15	.15	.17	.17	.18	.52	.56	.58	.63	.64	–		
23. perc_o_6	.42	.48	.50	.26	.30	.32	.13	.15	.15	.17	.15	.16	.16	.18	.18	.19	.55	.60	.62	.67	.68	.71	–	
24. perc_o_5	.43	.48	.50	.26	.30	.32	.14	.15	.15	.17	.15	.16	.16	.18	.19	.19	.55	.60	.62	.68	.68	.72	.76	–
Experience IR with others via network																								
25. internwo10_14_2	.07	.08	.08	.06	.07	.08	.08	.09	.09	.11	.10	.10	.10	.12	.12	.13	.04	.04	.04	.04	.05	.05	.05	.05
26. internwo10_14_1	.06	.07	.08	.06	.07	.07	.08	.09	.09	.10	.10	.10	.10	.11	.12	.12	.04	.04	.04	.04	.04	.05	.05	.05
27. internwo10_14_3	.06	.07	.07	.05	.06	.07	.07	.08	.08	.09	.09	.09	.09	.10	.11	.11	.03	.03	.04	.04	.04	.04	.04	.04
Experience IR with others in physical proximity																								
28. wothers10_14_2	.11	.13	.13	.10	.12	.13	.07	.07	.08	.09	.05	.06	.06	.07	.07	.07	.07	.07	.08	.08	.08	.09	.09	.09
29. wothers10_14_3	.11	.12	.12	.10	.11	.12	.06	.07	.07	.08	.05	.05	.05	.06	.06	.06	.06	.07	.07	.08	.08	.08	.09	.09
30. wothers10_14_1	.10	.11	.12	.09	.11	.12	.06	.07	.07	.08	.05	.05	.05	.06	.06	.06	.06	.07	.07	.07	.08	.08	.08	.08
Experience IR on one's own and alone																								
31. allone10_14_2	.10	.11	.12	.08	.10	.10	.06	.06	.06	.07	.06	.06	.06	.07	.07	.07	.07	.08	.08	.09	.09	.09	.10	.10
32. allone10_14_1	.09	.10	.10	.07	.08	.09	.05	.06	.06	.07	.05	.05	.05	.06	.06	.06	.06	.07	.07	.08	.08	.08	.09	.09
33. allone10_14_3	.07	.08	.08	.06	.07	.07	.04	.04	.04	.05	.04	.04	.04	.05	.05	.05	.05	.05	.06	.06	.06	.06	.07	.07
Watch IR action passively																								
34. watch10_14_1	.11	.13	.13	.06	.07	.07	.04	.04	.04	.05	.04	.04	.05	.05	.05	.05	.09	.10	.10	.11	.11	.11	.12	.12
35. watch10_14_2	.12	.14	.14	.06	.07	.08	.04	.04	.04	.05	.05	.05	.05	.06	.06	.06	.09	.10	.10	.11	.11	.12	.13	.13
Performance																								
36. esllevel	.01	.01	.01	.00	.00	.00	.00	.00	.00	.00	.05	.05	.05	.06	.06	.06	.01	.01	.01	.01	.01	.01	.01	.01
Cognitive absorption																								
37. cogabs_8	.23	.26	.27	.22	.25	.26	.10	.11	.11	.13	.09	.09	.10	.11	.11	.11	.17	.19	.19	.21	.21	.22	.24	.24

Continued on next page

Table D.51 Correlations, Means, and Standard Deviations I - ML (untransformed data) – *Continued*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
38. cogabs_2	.23	.26	.27	.21	.25	.26	.10	.11	.11	.13	.09	.09	.09	.11	.11	.11	.17	.19	.19	.21	.21	.22	.24	.24
39. cogabs_1	.23	.26	.27	.22	.25	.26	.10	.11	.11	.13	.09	.09	.09	.11	.11	.11	.17	.19	.19	.21	.21	.22	.24	.24
40. cogabs_3	.20	.23	.24	.19	.21	.23	.09	.10	.10	.11	.08	.08	.08	.09	.10	.10	.15	.16	.17	.18	.18	.19	.21	.21
41. cogabs_5	.19	.22	.23	.18	.20	.22	.08	.09	.09	.11	.08	.08	.08	.09	.09	.09	.14	.16	.16	.17	.18	.19	.20	.20
42. cogabs_4	.19	.21	.22	.18	.20	.22	.08	.09	.09	.11	.07	.08	.08	.09	.09	.09	.14	.15	.16	.17	.17	.18	.19	.20
Institutional achievement feedback																								
43. clever_2	.21	.24	.25	.24	.27	.29	.11	.12	.12	.14	.09	.09	.09	.10	.10	.11	.06	.07	.07	.08	.08	.08	.09	.09
44. clever_1	.20	.22	.23	.22	.25	.27	.10	.11	.11	.13	.08	.08	.08	.10	.10	.10	.06	.06	.07	.07	.07	.08	.08	.08
45. clever_3	.18	.20	.21	.20	.23	.24	.09	.10	.10	.11	.07	.07	.07	.08	.09	.09	.05	.06	.06	.06	.06	.07	.07	.07
Quotidian achievement feedback																								
46. clever_4	.13	.15	.15	.14	.16	.17	.07	.07	.08	.09	.06	.06	.06	.07	.07	.08	.03	.04	.04	.04	.04	.04	.04	.05
47. clever_6	.13	.15	.15	.13	.15	.17	.07	.07	.07	.08	.06	.06	.06	.07	.07	.07	.03	.03	.04	.04	.04	.04	.04	.04
48. clever_7	.10	.11	.11	.10	.11	.12	.05	.05	.06	.06	.04	.04	.05	.05	.05	.05	.02	.03	.03	.03	.03	.03	.03	.03
Means and standard deviation (available data points only)																								
M	4.87	4.88	5.00	4.84	5.32	5.27	4.70	4.65	5.04	4.92	4.39	4.13	4.70	3.94	4.29	4.06	4.55	5.12	5.01	5.01	4.21	4.83	4.61	4.85
SD	1.34	1.39	1.32	1.45	1.36	1.39	1.42	1.47	1.43	1.42	1.52	1.63	1.68	1.77	1.73	1.78	1.67	1.49	1.47	1.45	1.62	1.51	1.49	1.49

Table D.52: Intercorrelations, Means, and Standard Deviations of Variables Included in the Analysis II - ML estimation with original performance data

Variable	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Experience IR with others via network																								
25. internwo10_14_2	–																							
26. internwo10_14_1	.88	–																						
27. internwo10_14_3	.80	.77	–																					
Experience IR with others in physical proximity																								
28. wothers10_14_2	.22	.22	.20	–																				
29. wothers10_14_3	.21	.20	.18	.71	–																			
30. wothers10_14_1	.20	.19	.18	.68	.63	–																		
Experience IR on one's own and alone																								
31. allone10_14_2	.28	.27	.25	.34	.32	.30	–																	
32. allone10_14_1	.25	.24	.22	.30	.28	.27	.76	–																
33. allone10_14_3	.20	.19	.17	.24	.22	.21	.59	.53	–															
Watch IR action passively																								
34. watch10_14_1	.05	.05	.04	.11	.10	.09	.10	.09	.07	–														
35. watch10_14_2	.05	.05	.05	.11	.11	.10	.11	.10	.08	.65	–													
Performance																								
36. esltrans	.02	.02	.02	-.02	-.02	-.02	-.02	-.01	-.01	.00	.00	–												
Cognitive absorption																								
37. cogabs_8	.13	.12	.11	.23	.21	.20	.24	.21	.17	.07	.08	-.01	–											
38. cogabs_2	.13	.12	.11	.23	.21	.20	.24	.21	.16	.07	.08	-.01	.67	–										
39. cogabs_1	.13	.12	.11	.23	.21	.20	.24	.21	.17	.07	.08	-.01	.67	.67	–									
40. cogabs_3	.11	.10	.10	.20	.18	.18	.21	.18	.14	.06	.07	-.01	.58	.58	.58	–								
41. cogabs_5	.10	.10	.09	.19	.18	.17	.20	.18	.14	.06	.06	-.01	.56	.55	.56	.48	–							
42. cogabs_4	.10	.10	.09	.19	.17	.17	.20	.17	.14	.06	.06	-.01	.55	.55	.55	.48	.45	–						
Institutional achievement feedback																								
43. clever_2	.13	.13	.12	.19	.18	.17	.12	.11	.08	.07	.08	-.02	.22	.22	.22	.19	.18	.18	–					
44. clever_1	.12	.12	.11	.18	.17	.16	.11	.10	.08	.07	.07	-.02	.21	.20	.21	.18	.17	.17	.66	–				

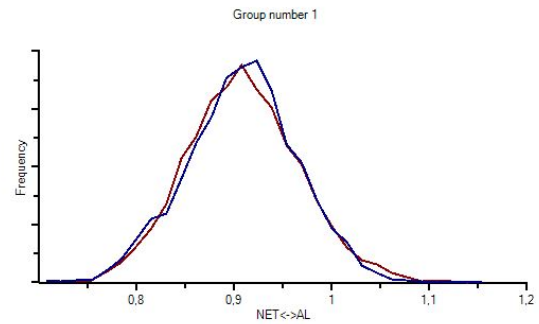
Continued on next page

Table D.52 Correlations, Means, and Standard Deviations II - ML (untransformed data) – *Continued*

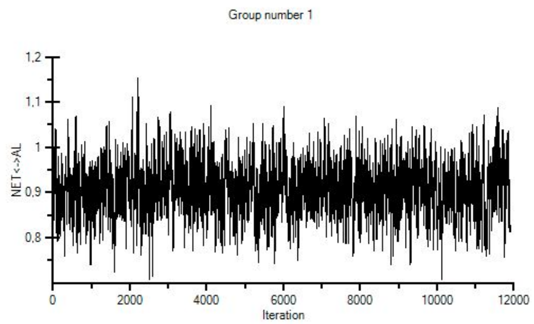
Variable	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
45. clever_3	.11	.11	.10	.16	.15	.14	.10	.09	.07	.06	.06	-.02	.18	.18	.18	.16	.15	.15	.59	.55	–			
Quotidian achievement feedback																								
46. clever_4	.17	.16	.15	.10	.09	.09	.08	.07	.06	.07	.07	.01	.10	.10	.10	.09	.08	.08	.46	.43	.38	–		
47. clever_6	.17	.16	.15	.10	.09	.09	.08	.07	.06	.07	.07	.01	.10	.10	.10	.08	.08	.08	.45	.42	.38	.72	–	
48. clever_7	.12	.12	.11	.07	.07	.06	.06	.05	.04	.05	.05	.01	.07	.07	.07	.06	.06	.06	.34	.31	.28	.53	.52	–
Means and standard deviation (available data points only)																								
M	4.29	4.28	4.12	5.08	5.13	5.01	4.69	4.75	3.63	3.73	3.90	31.55	5.47	5.27	5.60	4.98	5.19	4.95	5.42	5.12	5.57	4.52	4.18	4.08
SD	1.96	1.92	1.99	1.52	1.58	1.48	1.63	1.64	1.80	1.68	1.75	34.44	1.45	1.53	1.36	1.62	1.43	1.57	1.31	1.36	1.28	1.49	1.57	1.71

D.3.2 *Bayesian Analysis*

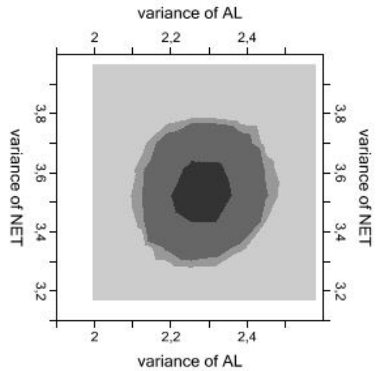
The following four diagnostic plots display convergence information concerning randomly picked relationships of the CIRME construct; more precisely, Figure D.1a shows that the posterior distributions of the parameter of the covariance between experiencing IR on one's own and alone (AL) and experiencing IR with others via network (NET); one can see that the distributions are almost identical for the first and the last thirds of the analysis samples which suggests that AMOS has correctly identified the important features of the posterior distribution of this covariance. Furthermore, Figure D.1b shows the sampled values of the parameter of the covariance just mentioned over time and thus how quickly the procedure converged for this parameter. In contrast, Figure D.1c and Figure D.1d both show bivariate marginal posterior distributions of estimands, where the former displays a two-dimensional plot of the bivariate posterior density of the variances of AL and NET (the three shades of gray represent 50%, 90%, and 95% credible regions, respectively) and the latter displays a three-dimensional surface plot of the marginal posterior distribution of their variances (all guidelines of how to read these plots are taken from Arbuckle, 2013).



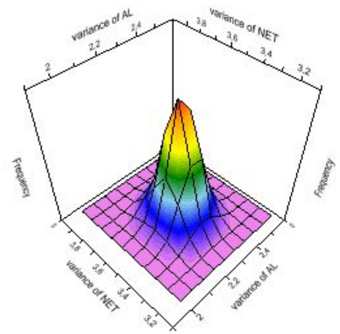
(a) Frequency polygon



(b) Time-series plot



(c) Posterior density



(d) Surface plot

Figure D.1: Examples of diagnostic plots for several CIRME parameter estimates

E PROGRAMMING RESOURCES

E.1 PACKAGE STRUCTURE AND JAVA LIBRARIES OF WEB CRAWLER

- ◯ OperatornameCrawler
  - ▽ trunk/src
    - ⋄ operatorname
      - ⋄ crawler
      - ⋄ db
      - ⋄ exceptions
      - ⋄ norm
      - ⋄ parser
      - ⋄ resourcehandling
      - ⋄ spss
      - Starter.java
  - ▽ trunk/test
    - ⋄ operatorname
      - ⋄ crawler
      - ⋄ ...
      - ⋄ spss
      - StarterTest.java
      - LoggerTest.java
  - □ JRE System Library [jre7]
  - □ Referenced Libraries
    - ◁ apache-log4j-extras-1.1.jar
    - ◁ log4j-1.2.16.jar
    - ◁ ant.jar
    - ◁ jericho-html-3.2.jar
    - ◁ commons-logging-1.1.1.jar
    - ◁ junit-4.10.jar
    - ◁ mysql-connector-java-5.1.18-bin.jar
    - ◁ slf4j-api-1.6.1.jar
  - ▽ trunk
    - ∘ analysis
    - ∘ config
    - ∘ etc
    - ∘ lib
    - ∘ log
    - ∘ mergefiles
    - ∘ sqlfiles
    - ∘ storedurlfiles
    - ∘ testfiles
    - ⁎ build.xml

Legend:
◯ = Project folder, ▽ = Source code folder, ⋄ = Java package, □ = Library container,
◁ = Jar file, ∘ = Folder.

E.2 AMOS PROGRAM CODE

Listing 1: Visual Basic Code for ML-CFA with FIML in AMOS 22. Adapted from Allison, P. D. (2003): Missing Data Techniques for Structural Equation Modeling. Journal of Abnormal Psychology, 112(4), p. 557, and Arbuckle, J. L. (2012): Amos 21 user's guide.

```
#Region "Header"
Imports System
Imports System.Diagnostics
Imports Microsoft.VisualBasic
Imports AmosEngineLib
Imports AmosGraphics
Imports AmosEngineLib.AmosEngine.TMatrixID
Imports PBayes
Imports MiscAmosTypes
Imports MiscAmosTypes.cDatabaseFormat
#End Region
Module MainModule
'When ModelMeansAndIntercepts method is used, following default
    assumptions are made about means that are not constrained or
    fixed at constant values by use of the MStructure or Mean
    methods:
'—The means of observed, exogenous variables (ie indicators) are
    free parameters.
'—The means of unobserved, exogenous variables (ie error terms
    and latent variables) are fixed at zero.
'Remarks: If parameterValue and parameterName are omitted, the
    mean is an unconstrained parameter.
Sub Main()
  Dim Sem As New AmosEngine
     Try
        Sem.TextOutput()
        Sem.Standardized()
        Sem.Smc()
        Sem.ModelMeansAndIntercepts()
        Sem.BeginGroup("E:\...\filename.sav")

        'Include means into estimation; values are set
            automatically.
        Sem.Mean("watch10_14_1")
        Sem.Mean("watch10_14_2")
        Sem.Mean("watch10_14_3")
        Sem.Mean("allone10_14_1")
        Sem.Mean("allone10_14_2")
        Sem.Mean("allone10_14_3")
        Sem.Mean("wothers10_14_1")
        Sem.Mean("wothers10_14_2")
        Sem.Mean("wothers10_14_3")
        Sem.Mean("internwo10_14_1")
        Sem.Mean("internwo10_14_2")
        Sem.Mean("internwo10_14_3")
        Sem.Mean("age")

        'Define relationships between latent variables, their
            individual indicators, and their respective error
            terms.
        Sem.AStructure("watch10_14_1 = watch + (1)e1")
        Sem.AStructure("watch10_14_2 = watch + (1)e2")
        Sem.AStructure("watch10_14_3 = (1)watch + (1)e3")
        Sem.AStructure("allone10_14_1 = allone + (1)e4")
        Sem.AStructure("allone10_14_2 = allone + (1)e5")
        Sem.AStructure("allone10_14_3 = (1)allone + (1)e6")
        Sem.AStructure("wothers10_14_1 = with_others + (1)e7")
```

```
        Sem.AStructure("wothers10_14_2 = with_others + (1)e8")
        Sem.AStructure("wothers10_14_3 = (1)with_others + (1)e9")
        Sem.AStructure("internwo10_14_1 = network + (1)e10")
        Sem.AStructure("internwo10_14_2 = network + (1)e11")
        Sem.AStructure("internwo10_14_3 = (1)network + (1)e12")

        'To include the auxiliary variable "age", the model must
            allow for correlations of age with
        '(1) all measured exogenous variables (not applicable
            here, age is the only one) and with
        '(2) the error terms for each measured endogenous
            variable, Allison (2003), p. 550.
        Sem.AStructure("age <> e1")
        Sem.AStructure("age <> e2")
        Sem.AStructure("age <> e3")
        Sem.AStructure("age <> e4")
        Sem.AStructure("age <> e5")
        Sem.AStructure("age <> e6")
        Sem.AStructure("age <> e7")
        Sem.AStructure("age <> e8")
        Sem.AStructure("age <> e9")
        Sem.AStructure("age <> e10")
        Sem.AStructure("age <> e11")
        Sem.AStructure("age <> e12")

        'Suppress estimation of these covariances.
        Sem.Cov("age", "watch", 0)
        Sem.Cov("age", "allone", 0)
        Sem.Cov("age", "with_others", 0)
        Sem.Cov("age", "network", 0)
        Sem.FitModel
    Finally
        Sem.Dispose
  End Try
 End Sub
End Module
```

Listing 2: Code required to fit the saturated and independent model when using the VB program editor of AMOS (code for AMOS 4). Adapted from http://www.amosdevelopment.com/support/tips/basic_allfitmeasures.htm.

```
Sub Main()
        Const FileName As String = "c:\...\grant_x.sav"

    Dim SaturatedCmin As Double
    Dim LSaturatedResult As Long
    Dim IndependenceCmin As Double
    Dim LIndependenceResult As Long

        'Try to fit the saturated and independence models
        LSaturatedResult = SaturatedOrIndependence(SaturatedCmin,
             False, 6, _
                        Array("visperc", "cubes", "lozenges", "
                            paragrap", "sentence", "wordmean"), _
                        FileName)
        LIndependenceResult = SaturatedOrIndependence(
            IndependenceCmin, True, 6, _
                        Array("visperc", "cubes", "lozenges", "
                            paragrap", "sentence", "wordmean"), _
                        FileName)

        'Create a new instance of the Amos engine
    Dim Sem As AmosEngine
    Set Sem = New AmosEngine
```

```
        'Tell the Amos engine about the fit of the saturated and
            independence models
        Sem.SetSaturatedFit LSaturatedResult = 0, SaturatedCmin
        Sem.SetIndependenceFit LIndependenceResult = 0,
            IndependenceCmin
        ...
End Sub

...

Function SaturatedOrIndependence(CMin As Double, Independence As
    Boolean, _
        NObservedVariables As Long, ObservedVariables_0b As
            Variant, _
        FileName As String, Optional TableName As String, _
        Optional GroupingVariable As String, _
        Optional GroupValue As Variant) As Long
    Dim Sem As AmosEngine
    Dim i As Integer
    Dim STemp As String
    SaturatedOrIndependence = 0
    On Error GoTo EHandler
    Set Sem = New AmosEngine
    Sem.ModelMeansAndIntercepts

    Call Sem.GenerateDefaultCovariances(Not Independence)

    Sem.BeginGroup FileName, TableName, GroupingVariable,
        GroupValue
    For i = 0 To NObservedVariables - 1
        STemp = ObservedVariables_0b(i)
        Sem.Mean STemp
    Next

    If Sem.FitModel() = 0 Then
        CMin = Sem.CMin
    Else
        SaturatedOrIndependence = 1
    End If
Quit:
    Set Sem = Nothing
    DoEvents
    Exit Function

EHandler:
    SaturatedOrIndependence = 1
    GoTo Quit
End Function
```

E.3 R SOURCE CODE

Listing 3: Minimal sample size calculation given null and alternative RMSEA values for a close fit, degrees of freedom, as well as alpha and power levels. Adapted from Preacher, K. J., & Coffman, D. L. (2006, May): Computing Power and Minimum Sample Size for RMSEA (Computer software), available from http://quantpsy.org/.

```
#computation of minimum sample size for test of fit
rmsea0 <- 0.05 #null hypothesized RMSEA
rmseaa <- 0.08 #alternative hypothesized RMSEA
d <- 14 #degrees of freedom
alpha <- 0.01 #alpha level
desired <- 0.99 #desired power

#initialize values
pow <- 0.0
n <- 0
#begin loop for finding initial level of n
while (pow<desired) {
  n <- n+100
  ncp0 <- (n-1)*d*rmsea0^2
  ncpa <- (n-1)*d*rmseaa^2
  #compute power
  if(rmsea0<rmseaa) {
      cval <- qchisq(alpha,d,ncp=ncp0,lower.tail=F)
      pow <- pchisq(cval,d,ncp=ncpa,lower.tail=F)
  }
  else {
      cval <- qchisq(1-alpha,d,ncp=ncp0,lower.tail=F)
      pow <- 1-pchisq(cval,d,ncp=ncpa,lower.tail=F)
  }
}

#begin loop for interval halving
foo <- -1
newn <- n
interval <- 200
powdiff <- pow - desired
while (powdiff>.001) {
  interval <- interval*.5
  newn <- newn + foo*interval*.5
  ncp0 <- (newn-1)*d*rmsea0^2
  ncpa <- (newn-1)*d*rmseaa^2
  #compute power
  if(rmsea0<rmseaa) {
      cval <- qchisq(alpha,d,ncp=ncp0,lower.tail=F)
      pow <- pchisq(cval,d,ncp=ncpa,lower.tail=F)
  }
  else {
      cval <- qchisq(1-alpha,d,ncp=ncp0,lower.tail=F)
      pow <- 1-pchisq(cval,d,ncp=ncpa,lower.tail=F)
  }
  powdiff <- abs(pow-desired)
  if (pow<desired) {
    foo <- 1
  }
  if (pow>desired) {
    foo <- -1
  }
}

minn <- newn
print(minn)
```

Listing 4: Box-Cox Transformation. Adapted from Wessa P., (2013), Box-Cox Normality Plot (v1.1.5) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL: http://www.wessa.net/rwasp_boxcoxnorm.wasp/. R code based on: NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 2006-10-03.

```
par2 <- abs(as.numeric(par2)*100)
par3 <- as.numeric(par3)*100
if(par4=="") par4 <- 0
par4 <- as.numeric(par4)
numlam <- par2 + par3 + 1
x <- x + par4
n <- length(x)
c <- array(NA,dim=c(numlam))
l <- array(NA,dim=c(numlam))
mx <- -1
mxli <- -999

for (i in 1:numlam)      {
        l[i] <- (i-par2-1)/100
        if (l[i] != 0)   {
                if (par1 == "Full Box-Cox transform") x1 <- (x^l[
                    i] - 1) / l[i]
                if (par1 == "Simple Box-Cox transform") x1 <- x^l
                    [i]
        }
        else {
                x1 <- log(x)
        }
        c[i] <- cor(qnorm(ppoints(x), mean=0, sd=1),x1)
        if (mx < c[i])   {
                mx <- c[i]
                mxli <- l[i]
                x1.best <- x1
                }
}

c
mx
mxli
x1.best

if (mxli != 0)   {
        if (par1 == "Full Box-Cox transform") x1 <- (x^mxli - 1)
            / mxli
        if (par1 == "Simple Box-Cox transform") x1 <- x^mxli
        }
        else {
                x1 <- log(x)
}

bitmap(file="test1.png")
plot(l,c,main="Box-Cox Normality Plot", xlab="Lambda",ylab="
    correlation")
mtext(paste("Optimal Lambda =",mxli))
grid()
dev.off()
bitmap(file="test2.png")
hist(x,main="Histogram of Original Data",xlab="X",ylab="frequency
    ")
grid()
dev.off()
bitmap(file="test3.png")
```

```
hist(x1,main="Histogram of Transformed Data", xlab="X",ylab="
    frequency")
grid()
dev.off()
bitmap(file="test4.png")
qqnorm(x)
qqline(x)
grid()
mtext("Original Data")
dev.off()
bitmap(file="test5.png")
qqnorm(x1)
qqline(x1)
grid()
mtext("Transformed Data")
dev.off()
load(file="createtable")
a<-table.start()
a<-table.row.start(a)
a<-table.element(a,"Box-Cox Normality Plot",2,TRUE)
a<-table.row.end(a)
a<-table.row.start(a)
a<-table.element(a,"# observations x",header=TRUE)
a<-table.element(a,n)
a<-table.row.end(a)
a<-table.row.start(a)
a<-table.element(a,"maximum correlation",header=TRUE)
a<-table.element(a,mx)
a<-table.row.end(a)
a<-table.row.start(a)
a<-table.element(a,"optimal lambda",header=TRUE)
a<-table.element(a,mxli)
a<-table.row.end(a)

if(mx<0) {
        a<-table.row.start(a)
        a<-table.element(a,"Warning: maximum correlation is
            negative! The Box-Cox transformation must not be used
            .",2)
        a<-table.row.end(a)
}
a<-table.end(a)
table.save(a,file="mytable.tab")

if(par5=="Yes") {
        a<-table.start()
        a<-table.row.start(a)
        a<-table.element(a,"Obs.",header=T)
        a<-table.element(a,"Original",header=T)
        a<-table.element(a,"Transformed",header=T)
        a<-table.row.end(a)
        for (i in 1:n) {
                a<-table.row.start(a)
                a<-table.element(a,i)
                a<-table.element(a,x[i])
                a<-table.element(a,x1.best[i])
                a<-table.row.end(a)
        }
a<-table.end(a)
table.save(a,file="mytable1.tab")
}
```

REFERENCES

Ackerman, P. L. & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*(2), 219–245. doi:10.1037/0033-2909.121.2.219

Ackerman, P. L. & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, *13*(2), 57–78. doi:10.1037/1076-898X.13.2.57

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, *16*(2), 227–247.

Agarwal, R., Animesh, A., & Prasad, K. (2009). Social interactions and the "digital divide": Explaining variations in Internet use. *Information Systems Research*, *20*(2), 277–294. doi:10.1287/isre.1080.0194

Agarwal, R. & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, *24*(4), 665–694.

Agarwal, R. & Lucas, H. C. J. (2005). The Information Systems identity crisis: Focusing on high-visibility and high-impact research. *MIS Quarterly*, *29*(3), 381–398.

Agarwal, R., Sambamurthy, V., & Stair, R. M. (1997). Cognitive absorption and the adoption of new information technologies. *Academy of Management Best Papers Proceedings*, *8*(1), 293–297. doi:10.5465/AMBPP.1997.4983719

Aguirre-Urreta, M. I. & Marakas, G. M. (2012). Revisiting bias due to construct misspecification: Different results from considering coefficients in standardized form [Research note]. *MIS Quarterly*, *36*(1), 123–138.

Albors, J., Ramos, J., & Hervas, J. (2008). New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source. *International Journal of Information Management*, *28*(3), 194–202. doi:10.1016/j.ijinfomgt.2007.09.006

Allen, G. N., Ball, N., & Smith, H. J. (2011). Information Systems research behaviors: What are the normative standards? *MIS Quarterly*, *35*(3), 533–551.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545–557. doi:10.1037/0021-843X.112.4.545

Allison, P. D. (2009). Missing Data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology*. London: Sage Publications.

Allison, P. D. (2012). Handling missing data by maximum likelihood [Keynote address]. In A. T. Kuligowski (Ed.), *Proceedings of the 2012 SAS Global Forum.* Statistics and Data Analysis.

Allport, C. D. & Kerler, W. A. I. (2003). A research note regarding the development of the consensus on appropriation scale [Research note]. *Information Systems Research*, *14*(4), 356–359.

Alsina-Jurnet, I. & Gutiérrez-Maldonado, J. (2010). Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments. *International Journal of Human-Computer Studies*, *68*(10), 788–801. doi:10.1016/j.ijhcs.2010.07.001

Ames, D. R. & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior*, *28*(3). doi:10.1023/B:JONB.0000039649.20015.0e

Anderson, D. R. & Hanson, K. G. (2010). From blooming, buzzing confusion to media literacy: The early development of television viewing. *Developmental Review*, *30*(2), 239–255. doi:10.1016/j.dr.2010.03.004

Anderson, J. C. & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. doi:10.1037/0033-2909.103.3.411

Andridge, R. R. & Little, R. J. A. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, *78*(1), 40–64. doi:10.1111/j.1751-5823.2010.00103.x

Animesh, A., Pinsonneault, A., Yang, S.-B., & Oh, W. (2011). An odyssey into virtual worlds: Exploring the impacts of technological and spatial environments [Special issue]. *MIS Quarterly*, *35*(3), 789–810.

Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology*, *14*(2), 105–112. doi:10.1037/a0018985

Arakji, R. Y. & Lang, K. R. (2008). Avatar business value analysis: A method for the evaluation of business value creation in virtual commerce. *Journal of Electronic Commerce Research*, *9*(3), 207–218.

Aral, S., Brynjolfsson, E., & Alstyne, M. v. (2012). Information, technology, and information worker productivity. *Information Systems Research*, *23*(3-part-2), 849–867. doi:10.1287/isre.1110.0408

Arbuckle, J. L. (2013). *Amos 22 user's guide.* IBM Corporation.

Armstrong, D. J. & Hardgrave, B. C. (2007). Understanding mindshift learning: The transition to object-oriented development. *MIS Quarterly*, *31*(3), 453–474.

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*(4), 441–452. doi:10.1002/job.320

Asparouhov, T. & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. doi:10.1080/10705510903008204

Aufderheide, P. & Firestone, C. M. (1993). *Media literacy: A report of the National Leadership Conference on Media Literacy, the Aspen Institute Wye Center, Queenstown Maryland, December 7-9, 1992*. Communications and society. Washington, D.C: Communications and Society Program, the Aspen Institute.

Austin, E. J., Saklofske, D. H., Huang, S. H., & McKenney, D. (2004). Measurement of trait emotional intelligence: Testing and cross-validating a modified version of Schutte et al.'s (1998) measure. *Personality and Individual Differences*, *36*(3), 555–562. doi:10.1016/S0191-8869(03)00114-4

Backhaus, K., Erichson, B., & Weiber, R. (2011). *Fortgeschrittene multivariate Analysemethoden: Eine anwendungsorientiere Einführung*. Berlin: Springer.

Bagozzi, R. P. (2007). The legacy of the Technology Acceptance Model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, *8*(4), 244–254.

Bagozzi, R. P. (2011). Measurement and meaning in Information Systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly*, *35*(2), 261–292.

Bagozzi, R. P. & Dholakia, U. M. (2006). Antecedents and purchase consequences of customer participation in small group brand communities. *International Journal of Research in Marketing*, *23*(1), 45–61. doi:10.1016/j.ijresmar.2006.01.005

Bagozzi, R. P. & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, *1*(1), 45–87. doi:10.1177/109442819800100104

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, *27*(2), 184–206. doi:10.1177/0092070399272005

Bagozzi, R. P. & Moore, D. J. (1994). Public service advertisements: Emotions and empathy guide prosocial behavior. *The Journal of Marketing*, *58*(1), 56–70.

Bagozzi, R. P. & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1).

Bagozzi, R. P. & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, *40*(1), 8–34. doi:10.1007/s11747-011-0278-x

Bailenson, J. N., Yee, N., Brave, S., Merget, D., & Koslow, D. (2007). Virtual interpersonal touch: Expressing and recognizing emotions through haptic devices. *Human-Computer Interaction*, *22*, 325–353.

Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*, *317*(5837), 472–476. doi:10.1126/science.1146930

Bainbridge, W. S. (Ed.). (2010). *Online worlds: Convergence of the real and the virtual.* London: Springer. doi:10.1007/978-1-84882-825-4_20

Bakker, R. (2009). Re-measuring left–right: A comparison of SEM and Bayesian measurement models for extracting left–right party placements. *Electoral Studies*, *28*(3), 413–421. doi:10.1016/j.electstud.2009.05.008

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287. doi:10.1016/0749-5978(91)90022-L

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology*, *3*(3), 265–299. doi:10.1207/S1532785XMEP0303_03

Baraldi, A. N. & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37. doi:10.1016/j.jsp.2009.10.001

Barki, H., Titah, R., & Boffo, C. (2007). Information System use–related activity: An expanded behavioral conceptualization of individual-level Information System use. *Information Systems Research*, *18*(2), 173–192. doi:10.1287/isre.1070.0122

Barnett, J. & Coulson, M. (2010). Virtually real: A psychological perspective on massively multiplayer online games. *Review of General Psychology*, *14*(2), 167–179.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. doi:10.1037/0022-3514.51.6.1173

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824. doi:10.1016/j.paid.2006.09.018

Bartle, R. A. (2004). *Designing virtual worlds.* Indianapolis, Ind: New Riders. Retrieved from http://books.google.ie/books?id=z3VP7MYKqaIC

Baskerville, R. & Dulipovici, A. (2006). The theoretical foundations of knowledge management. *Knowledge Management Research & Practice*, *4*(2), 83–105. doi:10.1057/palgrave.kmrp.8500090

Basori, A. H., Daman, D., Sunar, M. S., & Bade, A. (2007). The potential of human haptic emotion as technique for virtual human characters movement to augment interactivity in virtual reality game. *The International Journal of Virtual Reality*, *7*(2), 27–32.

Bastian, V. A., Burns, N. R., & Nettelbeck, T. (2005). Emotional intelligence predicts life skills, but not as well as personality and cognitive abilities. *Personality and Individual Differences*, *39*(6), 1135–1145. doi:10.1016/j.paid.2005.04.006

Bateman, P. J., Gray, P. H., & Butler, B. S. (2010). The impact of community commitment on participation in online communities [Research note]. *Information Systems Research*. doi:10.1287/isre.1090.0265

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*(4), 425–430.

Beaudry, A. & Pinsonneault, A. (2010). The other side of acceptance: Studying the direct and indirect effects of emotions on information technology use. *MIS Quarterly*, *34*(4), 89–710.

Beauducel, A. & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(1), 41–75. doi:10.1207/s15328007sem1201_3

Becker, T. (2003). Is emotional intelligene a viable concept? *The Academy of Management Review*, *28*(2), 192–195.

Beier, M. E. & Ackerman, P. L. (2005). Working memory and intelligence: Different constructs. Reply to Oberauer et al. (2005) and Kane et al. (2005). *Psychological Bulletin*, *131*(1), 72–75. doi:10.1037/0033-2909.131.1.72

Bélanger, F., Collins, R. W., & Cheney, P. H. (2001). Technology requirements and work group communication for telecommuters. *Information Systems Research*, *12*(2), 155–176. doi:10.1287/isre.12.2.155.9695

Bell, B. S. & Kozlowski, W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology*, *87*(3), 497–505. doi:10.1037/0021-9010.87.3.497

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, *92*(3), 595–615. doi:10.1037/0021-9010.92.3.595

Benbasat, I. & Barki, H. (2007). Quo vadis, TAM? *Journal of the Association for Information Systems*, *8*(4), 212–218.

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*(1), 121–130. doi:10.1037/a0017767

Bentler, P. M. & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, 563–592. doi:10.1146/annurev.psych.47.1.563

Bentler, P. M. & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods*, *15*(2), 111–123.

Berente, N., Hansen, S., Pike, J. C., & Bateman, P. J. (2011). Arguing the value of virtual worlds: Patterns of discursive sensemaking of an innovative technology. *MIS Quarterly*, *35*(3), 685–709.

Bergami, M. & Bagozzi, R. P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology*, *39*(4), 555–577. doi:10.1348/014466600164633

Bernstein, I. H. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*(3), 467–477. doi:10.1037/0033-2909.105.3.467

Bers, M. U. (2010). Let the games begin: Civic playing on high-tech consoles. *Review of General Psychology*, *14*(2), 147–153. doi:10.1037/a0019490

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, *3*(2). doi:10.1111/j.1083-6101.1997.tb00070.x

Black, A. C., Harel, O., & McCoach, B. D. (2011). Missing data techniques for multilevel data: Implications of model misspecification. *Journal of Applied Statistics*, *38*(9), 1845–1865. doi:10.1080/02664763.2010.529882

Blalock, H. M. J. (1963). Making causal inferences for unmeasured variables from correlations among indicators. *American Journal of Sociology*, *69*(1), 53–62.

Blümel, E., Termath, W., & Haase, T. (2009). Virtual reality platforms for education and training in industry. *International Journal of Advanced Corporate Learning*, *2*(2). doi:10.3991/ijac.v2i2.870

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35*(2), 359–372.

Bollen, K. A. & Hoyle, R. H. (2012). Latent variables in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 56–67). New York: Guilford Press.

Boomsma, A., Hoyle, R. H., & Panter, A. T. (2012). The structural equation modeling research report. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 341–358). New York: Guilford Press.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6th ed.). Springer-Lehrbuch. Berlin: Springer.

Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2nd ed.). Berlin: Springer.

Bösche, W. (2009). Violent content enhances video game performance. *Journal of Media Psychology: Theories, Methods, and Applications*, *21*(4), 145–150. doi:10.1027/1864-1105.21.4.145

Boudreau, M.-C., Gefen, D., & Straub, D. W. (2001). Validation in Information Systems research: A state-of-the-art assessment. *MIS Quarterly*, *25*(1), 1–16.

Bovaird, J. A. & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 495–511). New York: Guilford Press.

Bowman, D. A., Sowndararajan, A., Ragan, E. D., & Kopper, R. (2009). Higher levels of immersion improve procedure memorization performance. In M. Hirose, D. Schmalstieg, C. A. Wingrave, & K. Nishimura (Eds.), *Proceedings of the 2009 Joint Virtual Reality Conference of EGVE - ICAT - EuroVR* (pp. 121–128).

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, *26*(2), 211–252.

Bracken, C. C. & Skalski, P. (2009). Telepresence and video games: The impact of image quality. *PsychNology Journal*, *7*(1), 101–112.

Brackett, M. A. & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, *29*(9), 1147–1158. doi:10.1177/0146167203254596

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology*, *91*(4), 780–795. doi:10.1037/0022-3514.91.4.780

Brandtzæg, P. B. (2010). Towards a unified media-user typology (MUT): A meta-analysis and review of the research literature on media-user typologies. *Computers in Human Behavior*, *26*(5), 940–956. doi:10.1016/j.chb.2010.02.008

Bray, D. A. & Konsynski, B. R. (2007). Virtual worlds: Multi-disciplinary research opportunities. *ACM SIGMIS Database*, *38*(4), 17. doi:10.1145/1314234.1314239

Brown, J. D. (2011). Likert items and scales of measurement? *SHIKEN: The JALT Testing & Evaluation SIG Newsletter*, *15*(1), 10–14.

Brown, T. A. & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). New York: Guilford Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures [Abstract]. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. doi:10.1111/j.2044-8317.1984.tb00789.x

Brunelle, E. & Lapierre, J. (2008). Testing media richness theory to explain consumers' intentions of buying online. In D. Fensel & H. Werthner (Eds.), *Proceedings of the Tenth International Conference on Electronic Commerce (ICEC 2008)*. New York: Association for Computing Machinery.

Buckler, F. & Hennig-Thurau, T. (2008). Identifying hidden structures in marketing's structural models through Universal Structure Modeling: An explorative Bayesian neural network complement to LIS-

REL and PLS. *Marketing : Journal of Research and Management*, *4*(2), 47–66.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3rd ed.). München: Pearson Studium.

Buijzen, M., Rozendaal, E., Moorman, M., & Tanis, M. (2008). Parent versus child reports of parental advertising mediation: Exploring the meaning of agreement. *Journal of Broadcasting & Electronic Media*, *52*(4), 509–525. doi:10.1080/08838150802437180

Burke, B. (2008). Media literacy in the digital age: Implications for scholars and students. *RCA Vestnik (Russian Communication Association)*, 14–24.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*(1), 60–77. doi:10.1037/0022-3514.90.1.60

Burton-Jones, A. & Straub, D. W. (2006). Reconceptualizing system usage: An approach and empirical test. *Information Systems Research*, *17*(3), 228–246.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, *12*(4), 346–362. doi:10.1287/isre.12.4.346.9703

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(2), 272–300. doi:10.1207/s15328007sem1102_8

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Cahalane, M., Feller, J., & Finnegan, P. (2012). Seeking the entanglement of immersion and emergence: Reflections from an analysis of the state of IS research on virtual worlds. In F. G. Joey (Ed.), *Proceedings of the Thirty-Third International Conference on Information Systems (ICIS 2012).*

Campbell, D. D., Rijsdijk, F. V., & Sham, P. C. (2007). Computation of individual latent variable scores from data with multiple missingness patterns. *Behavior Genetics*, *37*(2), 408–422. doi:10.1007/s10519-006-9123-2

Cao, Y., Glukhova, A., Klamma, R., Renzel, D., & Spaniol, M. (2008). Measuring community satisfaction across gaming communities. In Institute of Electrical and Electronics Engineers (Ed.), *Proceedings of the First IEEE International Conference on Ubi-Media Computing and Workshops (UMEDIA 2008).*

Caperton, I. H. (2010). Toward a theory of game-media literacy: Playing and building as reading and writing. *International Journal of Gaming and Computer-Mediated Simulations*, *2*(1), 1–16.

Carifio, J. & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116. doi:10.3844/jssp.2007.106.116

Carlson, P. J. & Davis, G. B. (1998). An investigation of media selection among directors and managers: From self to other orientation. *MIS Quarterly*, *22*(3), 335–362.

Carton, A. M. & Aiello, J. R. (2009). Control and anticipation of social interruptions: Reduced stress and improved task performance. *Journal of Applied Social Psychology*, *39*(1), 169–185.

Castronova, E. (2005). Real products in imaginary worlds. *Harvard Business Review*, 20–22.

Cattell, R. B. (1940). A culture-free intelligence test I. *Journal of Educational Psychology*, *31*(3), 161–179. doi:10.1037/h0059043

Cattell, R. B. (1973). *Manual for the Cattell Culture Fair Intelligence Tests*. Champaign, IL: Institute for Personality & Ability Testing. Retrieved from http://books.google.de/books?id=wSrhMQAACAAJ

Cenfetelli, R. T. & Bassellier, G. (2009). Interpretation of formative measurement in Information Systems research. *MIS Quarterly*, *33*(4), 689–707.

Chakraborty, I., Hu, P. J.-H., & Cui, D. (2008). Examining the effects of cognitive style in individuals' technology use decision making. *Decision Support Systems*, *45*(2), 228–241. doi:10.1016/j.dss.2007.02.003

Chaturvedi, A. R., Dolk, D. R., & Drnevich, P. L. (2011). Design principles for virtual worlds. *MIS Quarterly*, *35*(3), 673–684.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models [Teacher's corner]. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 471–492.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494. doi:10.1177/0049124108314720

Chin, W. W. (1995). Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, *2*, 315–319. Retrieved from http://disc-nt.cba.uh.edu/chin/technologystudies.pdf

Chin, W. W. (1998). Issues and opinion on structural equation modeling: Commentary. *MIS Quarterly*, *22*(1), vii–xvi.

Chin, W. W. & Lee, M. K. (2000). A proposed model and measurement instrument for the formation of IS satisfaction: The case of end-user computing satisfaction. In *Proceedings of the Twenty-First International Conference on Information Systems (ICIS 2000)*. Association for Information Systems.

Chin, W. W. & Marcolin, B. L. (2001). The future of diffusion research. *ACM SIGMIS Database*, *32*(3), 7. doi:10.1145/506724.506726

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research*, *14*(2), 189–217. doi:10.1287/isre.14.2.189.16018

Chin, W. W. & Todd, P. A. (1995). On the use, usefulness, and ease of use of structural equation modeling in MIS research: A note of caution. *MIS Quarterly*, *19*(2), 237–246.

Cho, C.-H. & Cheon, H. J. (2005). Children's exposure to negative Internet content: Effects of family context. *Journal of Broadcasting & Electronic Media*, *49*(4), 488–509. doi:10.1207/s15506878jobem4904_8

Choi, D. & Kim, J. (2004). Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contents. *CyberPsychology & Behavior*, *7*(1), 11–24. doi:10.1089/109493104322820066

Chou, C.-P. & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 37–55). Thousand Oaks: Sage Publications.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*(2), 347–357. doi:10.1111/j.2044-8317.1991.tb00966.x

Chou, C.-P. & Huh, J. (2012). Model modification in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 232–246). New York: Guilford Press.

Christ, W. G. (2004). Assessment, media literacy standards, and higher education. *American Behavioral Scientist*, *48*(1), 92–96. doi:10.1177/0002764204267254

Christakis, D. A. & Zimmerman, F. J. (2009). Young children and media: Limitations of current knowledge and future directions for research. *American Behavioral Scientist*, *52*(8), 1177–1185. doi:10.1177/0002764209331540

Churchill, G. A. J. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73.

Ciarrochi, J. V., Chan, A. Y., & Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality and Individual Differences*, *28*(3), 539–561. doi:10.1016/S0191-8869(99)00119-1

Clayes, E. L. & Anderson, A. H. (2007). Real faces and robot faces: The effects of representation on computer-mediated communication.

International Journal of Human-Computer Studies, *65*(6), 480–496. doi:10.1016/j.ijhcs.2006.10.005

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: Lawrence Erlbaum.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist*, *49*(12), 997–1003. doi:10.1037/0003-066X.49.12.997

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. doi:10.1037/1082-989X.6.4.330

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85*(5), 678–707. doi:10.1037//0021-9010.85.5.678

Compeau, D., Marcolin, B. L., Kelley, H., & Higgins, C. (2012). Generalizability of Information Systems research using student subjects–A reflection on our practices and recommendations for future research. *Information Systems Research*, *23*(4), 1093–1109. [Research commentary]. doi:10.1287/isre.1120.0423

Constant, D., Kiesler, S., & Sproull, L. (1994). What's mine is ours, or is it? A study of attitudes about information sharing. *Information Systems Research*, *5*(4), 400–421.

Conway, J. M. & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*(2), 147–168. doi:10.1177/1094428103251541

Cooper, A. & Petrides, K. V. (2010). A psychometric analysis of the Trait Emotional Intelligence Questionnaire–Short Form (TEIQue–SF) using item response theory. *Journal of Personality Assessment*, *92*(5), 449–457. doi:10.1080/00223891.2010.497426

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. doi:10.1037/0021-9010.78.1.98

Côté, S. & Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly*, *51*(1), 1–28.

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*(6), 885–901.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (1st ed.). New York: Harper & Row.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for Information Systems. *MIS Quarterly*, *11*(3), 354–366.

Dannecker, A., Dreßner, N., Richter, S., Fabisch, S., Lechner, U., & Ilsemann, A. (2008). Towards World of Warcraft as an experiment platform for teams. In *Proceedings of the Fourteenth Americas Conference on Information Systems (AMCIS 2008).*

Daus, C. S. & Ashkanasy, N. M. (2005). The case for the ability-based model of emotional intelligence in organizational behavior. *Journal of Organizational Behavior*, *26*(4), 453–466. doi:10.1002/job.321

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology*, *75*(4), 989–1015. doi:10.1037/0022-3514.75.4.989

Davis, A., Murphy, J., Owens, D., Khazanchi, D., & Zigurs, I. (2009). Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems*, *10*(2), 90–117.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, *50*(1), 61–77.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*(1), 1–38.

Dennis, A. R. (1996). Information exchange and use in group decision making: You can lead a group to information, but you can't make it think. *MIS Quarterly*, *20*(4), 433–457.

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly*, *32*(3), 575–600.

Dennis, A. R. & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research*, *9*(3).

Devaraj, S., Easley, R. F., & Crant, J. M. (2008). How does personality matter? Relating the Five-Factor model to technology acceptance and use [Research note]. *Information Systems Research*, *19*(1), 93–105. doi:10.1287/isre.1070.0153

Devonport, T. J. (2006). Perceptions of the contribution of psychology to success in elite kickboxing. *Journal of Sports Science and Medicine*, *5*(CSSI), 99–107.

Dewey, J. (1895). The theory of emotion. *Psychological Review*, *2*(1), 13–32. doi:10.1037/h0070927

Diamantopoulos, A. & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277.

Dickinger, A., Arami, M., & Meyer, D. (2008). The role of perceived enjoyment and social norm in the adoption of technology with network externalities. *European Journal of Information Systems*, *17*(1), 4–11. doi:10.1057/palgrave.ejis.3000726

Diefendorff, J. M. & Mehta, K. (2007). The relations of motivational traits with workplace deviance. *Journal of Applied Psychology*, *92*(4), 967–977. doi:10.1037/0021-9010.92.4.967

Do, C. B. & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*, *26*(8).

Donders, A. R. T., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*(10), 1087–1091. doi:10.1016/j.jclinepi.2006.01.014

Doty, D. H. & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods*, *1*(4), 374–406. doi:10.1177/109442819814002

Duan, W., Gu, B., & Whinston, A. B. (2009). Informational cascades and software adoption on the Internet: An empirical investigation. *MIS Quarterly*, *33*(1), 23–48.

Ducheneaut, N. & Moore, R. J. (2004). The social side of gaming: A study of interaction patterns in a massively multiplayer online game. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04).*

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). Building an MMO with mass appeal: A look at gameplay in World of Warcraft. *Games and Culture*, *1*(4), 281–317. doi:10.1177/1555412006292613

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2007). The life and death of online gaming communities: A look at guilds in World of Warcraft. In *Proceedings of the 2007 ACM Conference on Human Factors in Computing Systems (CHI '07).*

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3). doi:10.1111/1467-8721.01235

Dunson, D. B., Palomo, J., & Bollen, K. A. (2005). Bayesian structural equation modeling (Report No. 2005-5). In *Statistical and Applied Mathematical Sciences Institute Communications - Technical Reports*. Retrieved from http://www.samsi.info/communications/bayesian-structural-equation-modeling

Durcikova, A., Fadel, K. J., Butler, B. S., & Galletta, D. F. (2011). Knowledge exploration and exploitation: The impacts of psychological climate and knowledge management system access [Research note]. *Information Systems Research*, *22*(4), 855–866. doi:10.1287/isre.1100.0286

Edery, D. (2006). Reverse product placement in virtual worlds. *Harvard Business Review*, *84*(12), 24.

Edwards, M. C., Wirth, R. J., Houts, C. R., & Xi, N. (2012). Categorical data in the structural equation modeling framework. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 195–208). New York: Guilford Press.

Ehrlinger, J. & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*(1), 5–17. doi:10.1037/0022-3514.84.1.5

Elfenbein, H. A. & Ambady, N. (2002). Predicting workplace outcomes from the ability to eavesdrop on feelings. *Journal of Applied Psychology*, *87*(5), 963–971. doi:10.1037//0021-9010.87.5.963

Elfenbein, H. A., Foo, M. D., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior*, *31*(4), 205–223. doi:10.1007/s10919-007-0033-7

Ellis, R. & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, *91*(4), 451–471. doi:10.1348/000712600161934

Ellis, R., Tucker, M., Symes, E., & Vainio, L. (2007). Does selecting one visual object from several require inhibition of the actions associated with nonselected objects? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3).

Enders, C. K. (2001a). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(1), 128–141.

Enders, C. K. (2001b). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352–370. doi:10.1037/1082-989X.6.4.352

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, *8*(3), 322–337. doi:10.1037/1082-989X.8.3.322

Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. doi:10.1207/S15328007SEM0803_5

Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, *28*(5), 404–416. doi:10.1037/h0034679

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 56–83. doi:10.1080/10705519909540119

Ferrando, P. J., Anguiano-Carrasco, C., & Demestre, J. (2013). Combining IRT and SEM: A hybrid model for fitting responses and response certainties. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(2), 208–225. doi:10.1080/10705511.2013.769388

Fetscherin, M. & Lattemann, C. (2008). User acceptance of virtual worlds. *Journal of Electronic Commerce Research*, *9*(3), 231–242.

Fetscherin, M., Lattemann, C., & Lang, G. (2008). Virtual worlds research: A conceptual primer. *Journal of Electronic Commerce Research*, *9*(3), 192–194.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(2), 87–107. doi:10.1080/10705519709540063

Fiori, M. & Antonakis, J. (2012). Selective attention to emotional stimuli: What IQ and openness do, and emotional intelligence does not. *Intelligence*, *40*(3), 245–254. doi:10.1016/j.intell.2012.02.004

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 459–483). New York: Guilford Press.

Flanagan, D. P. & Harrison, P. L. (Eds.). (2012). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). New York: Guilford Press.

Floeck, F., Putzke, J., Steinfels, S., Fischbach, K., & Schoder, D. (2011). Imitation and quality of tags in social bookmarking systems: Collective intelligence leading to folksonomies. In T. Bastiaens, U. Baumöl, & B. Krämer (Eds.), *On collective intelligence* (Vol. 76, pp. 75–91). Advances in Intelligent and Soft Computing. Berlin: Springer. doi:10.1007/978-3-642-14481-3_7

Floyd, F. J. & Widaman, K. F. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment*, *7*(3), 286–299.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–1057. doi:10.1111/j.1467-9280.2007.02024.x

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. doi:10.2307/3151312

Fox, J., Byrnes, J. E., Boker, S., & Neale, M. C. (2012). Structural equation modeling in R with the sem and OpenMx Packages.

In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 325–340). New York: Guilford Press.

Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology*, *21*(1), 291–313.

Freudenthaler, H. H., Neubauer, A. C., Gabler, P., Scherl, W. G., & Rindermann, H. (2008). Testing and validating the Trait Emotional Intelligence Questionnaire (TEIQue) in a German-speaking sample. *Personality and Individual Differences*, *45*(7), 673–678. doi:10.1016/j.paid.2008.07.014

Freund, P. A. & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, *138*(2), 296–321.

Frith, C. D. & Frith, U. (2006). The neural basis of mentalizing [Minireview]. *Neuron*, *50*(4), 531–534. doi:10.1016/j.neuron.2006.05.001

Fritsch, T., Ritter, H., & Schiller, J. (2005). The effect of latency and network limitations on MMORPGs (A Field Study of Everquest2). In *Proceedings of the Fourth ACM SIGCOMM Workshop on Network and System Support for Games (NetGames 2005)*. New York, NY: Association for Computing Machinery. doi:10.1145/1103599.1103623

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, *14*(5), 449–468.

Frostling-Henningsson, M. (2009). First-person shooter games as a way of connecting to people: "Brothers in blood" [Rapid rommunication]. *CyberPsychology & Behavior*, *12*(5), 557–562. doi:10.1089/cpb.2008.0345

Fujioka, Y. & Austin, E. W. (2002). The relationship of family communication patterns to parental mediation styles. *Communication Research*, *29*(6), 642–665. doi:10.1177/009365002237830

Fullagar, C. J. & Kelloway, E. K. (2009). Flow at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology*, *82*(3), 595–615. doi:10.1348/096317908X357903

Furneaux, B. & Wade, M. (2011). An exploration of organizational level information systems discontinuance intentions. *MIS Quarterly*, *35*(3), 573–598.

Furnham, A. & Chamorro-Premuzic, T. (2004). Estimating one's own personality and intelligence scores. *British Journal of Psychology*, *95*(2), 149–160. doi:10.1348/000712604773952395

Furnham, A., Moutafi, J., & Chamorro-Premuzic, T. (2005). Personality and intelligence: Gender, the Big Five, self-estimated and psychometric intelligence. *International Journal of Selection and Assessment*, *13*(1), 11–24. doi:10.1111/j.0965-075X.2005.00296.x

Galanxhi, H. & Nah, F. F.-H. (2007). Deception in cyberspace: A comparison of text-only vs. avatar-supported medium. *International*

Journal of Human-Computer Studies, *65*(9), 770–783. doi:10.1016/j.ijhcs.2007.04.005

Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Nonnormality of data in structural equation models. *Transportation Research Record: Journal of the Transportation Research Board*, *2082*(-1), 116–124. doi:10.3141/2082-14

Garfield, M. J., Taylor, N. J., Dennis, A. R., & Satzinger, J. W. (2001). Modifying paradigms–Individual differences, creativity techniques, and exposure to ideas in group idea generation [Research report]. *Information Systems Research*, *12*(3), 322.

Gartner, Inc. (2007, April 24). Gartner says 80 percent of active Internet users will have a "Second Life" in the virtual world by the end of 2011. Stamford, Conneticut, USA. Retrieved from http://www.gartner.com/newsroom/id/503861

Gartner, Inc. (2008, May 15). Gartner says 90 percent of corporate virtual world projects fail within 18 months. Egham, UK. Retrieved from http://www.gartner.com/newsroom/id/670507

Gee, F. C., Browne, W. N., & Kawamura, K. (2005). Uncanny valley revisited. In *Proceedings of the Fourteenth IEEE International Workshop on Robots and Human Interactive Communication (RO-MAN 2005)* (pp. 151–157). Institute of Electrical and Electronics Engineers. doi:10.1109/ROMAN.2005.1513772

Gefen, D., Rigdon, E. E., & Straub, D. W. (2011). An update and extension to SEM guidelines for administrative and social science research [Editor's comments]. *MIS Quarterly*, *35*(2), iii–xiv.

Gelman, A. (2013). Understanding posterior p-values. *Electronic Journal of Statistics*, *7*, 2595–2602.

Gentile, D. A., Nathanson, A. I., Rasmussen, E. E., Reimer, R. A., & Walsh, D. A. (2012). Do you see what I see? Parent and child reports of parental monitoring of media. *Family Relations*, *61*(3), 470–487. doi:10.1111/j.1741-3729.2012.00709.x

Gignac, G. E. (2013). On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scales. *European Journal of Psychological Assessment*. doi:10.1027/1015-5759/a000181

Gignac, G. E., Palmer, B., Bates, T., & Stough, C. (2006). Differences in confirmatory factor analysis model close-fit index estimates obtained from AMOS 4.0 and AMOS 5.0 via full information maximum likelihood - no imputation: Corrections and extension to Palmer et al. (2003). *Australian Journal of Psychology*, *58*(3), 144–150. doi:10.1080/00049530500504088

Gill, G. & Bhattacherjee, A. (2009). Whom are we informing? Issues and recommendations for MIS research from an informing sciences perspective. *MIS Quarterly*, *33*(2), 217–235.

Gill, J. (2004). Introduction to the special issue. *Political Analysis*, *12*(4), 323–337. doi:10.1093/pan/mph021

Gintautas, V. & Hübler, A. (2007). Experimental evidence for mixed reality states in an interreality system. *Physical Review E*, *75*(5). doi:10.1103/PhysRevE.75.057201

Glockner Rist, A. & Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(4), 544–565. doi:10.1207/S15328007SEM1004_4

Goel, L., Johnson, N. A., Junglas, I., & Ives, B. (2011). From space to place: Predicting users' intentions to return to virtual worlds. *MIS Quarterly*, *35*(3), 749–771.

Goodhue, D. L. (2007). Comment on Benbasat and Barki's "Quo Vadis TAM" article [Commentary]. *Journal of the Association for Information Systems*, *8*(4), 220–222.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981–1001.

Goodhue, D. L. & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213–236.

Göritz, A. S. (2004). The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research*, *46*(3), 327–345.

Goyal, N. & Gomeni, R. (2013). A latent variable approach in simultaneous modeling of longitudinal and dropout data in schizophrenia trials. *European Neuropsychopharmacology*, *23*(11), 1570–1576. doi:10.1016/j.euroneuro.2013.03.004

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 80–100. doi:10.1207/S15328007SEM1001_4

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. doi:10.1146/annurev.psych.58.110405.085530

Graham, J. W. (2012a). Analysis of missing data. In J. W. Graham (Ed.), *Missing data* (pp. 47–69). New York: Springer. doi:10.1007/978-1-4614-4018-5_2

Graham, J. W. (2012b). Analysis with SPSS (versions without MI module) following multiple imputation with Norm 2.03. In J. W. Graham (Ed.), *Missing data* (pp. 95–109). New York: Springer. doi:10.1007/978-1-4614-4018-5_4

Graham, J. W. (Ed.). (2012c). *Missing data*. New York: Springer. doi:10.1007/978-1-4614-4018-5

Graham, J. W. (2012d). Multiple imputation and analysis with SAS. In J. W. Graham (Ed.), *Missing data* (pp. 151–190). New York: Springer. doi:10.1007/978-1-4614-4018-5_7

Graham, J. W. (2012e). Multiple imputation with Norm 2.03. In J. W. Graham (Ed.), *Missing data* (pp. 73–94). New York: Springer. doi:10.1007/978-1-4614-4018-5_3

Graham, J. W. & Coffman, D. L. (2012). Structural equation modeling with missing data. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 277–295). New York: Guilford Press.

Graham, J. W. & Collins, L. M. (2012). Using modern missing data methods with auxiliary variables to mitigate the effects of attrition on statistical power. In J. W. Graham (Ed.), *Missing data* (pp. 253–275). New York: Springer. doi:10.1007/978-1-4614-4018-5_11

Graham, J. W., Horn, M. L., & Taylor, B. J. (2012). Dealing with the problem of having too many variables in the imputation model. In J. W. Graham (Ed.), *Missing data* (pp. 213–228). New York: Springer. doi:10.1007/978-1-4614-4018-5_9

Green, S. B. & Thompson, M. S. (2012). A flexible structural equation modeling approach for analyzing means. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 393–416). New York: Guilford Press.

Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T., & Plomin, R. (2009). More than just IQ: School achievement is predicted by self-perceived abilities–But for genetic rather than environmental reasons. *Psychological Science*, *20*(6), 753–762. doi:10.1111/j.1467-9280.2009.02366.x

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. doi:10.1037/1082-989X.6.4.430

Grissom, R. J. & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, *6*(2), 135–146. doi:10.1037/1082-989X.6.2.135

Grittner, U., Gmel, G., Ripatti, S., Bloomfield, K., & Wicki, M. (2011). Missing value imputation in longitudinal measures of alcohol consumption. *International Journal of Methods in Psychiatric Research*, *20*(1), 50–61. doi:10.1002/mpr.330

Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., & Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, *67*(10), 842–849. doi:10.1016/j.ijhcs.2009.07.001

Grover, V., Straub, D. W., & Galluch, P. (2009). Turning the corner: The influence of positive thinking on the Information Systems field [Editor's comments]. *MIS Quarterly*, *33*(1), iii–viii.

Ha, S. & Stoel, L. (2009). Consumer e-shopping acceptance: Antecedents in a technology acceptance model. *Journal of Business Research*, *62*(5), 565–571.

Haefner, M. J. & Wartella, E. A. (1987). Effects of sibling coviewing on children's interpretations of television programs. *Journal of*

Broadcasting & Electronic Media, *31*(2), 153–168. doi:10.1080/08838158709386654

Haenlein, M. & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, *3*(4), 283–297. doi:10.1207/s15328031us0304_4

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *The Journal of Marketing Theory and Practice*, *19*(2), 139–152. doi:10.2753/MTP1069-6679190202

Halpern, S. D., Kohn, R., Dornbrand–Lo, A., Metkus, T., Asch, D. A., & Volpp, K. G. (2011). Lottery–based versus fixed incentives to increase clinicians' response to surveys. *Health Services Research*, *46*(5), 1663–1674. doi:10.1111/j.1475-6773.2011.01264.x

Hammervold, R. & Olsson, U. H. (2012). Testing structural equation models: The impact of error variances in the data generating process. *Quality & Quantity*, *46*(5), 1547–1570. doi:10.1007/s11135-011-9466-5

Hancock, G. R. & Liu, M. (2012). Bootstrapping standard errors and data–model fit statistics in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 296–306). New York: Guilford Press.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Heijden, H. v. d. (2004). User acceptance of hedonic Information Systems. *MIS Quarterly*, *28*(4), 695–704.

Heitjan, D. F. & Basu, S. (1996). Distinguishing "Missing at Random" and "Missing Completely at Random". *The American Statistician*, *50*(3), 207–213. doi:10.2307/2684656

Highsmith, J. A. (1978). Solving design problems more effectively. *MIS Quarterly*, *2*(4), 23–30.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121. doi:10.1177/109442819800100106

Hinkin, T. R. & Tracey, J. B. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, *2*(2), 175–186. doi:10.1177/109442819922004

Hirschheim, R. (2007). Introduction to the special issue on "Quo Vadis TAM–Issues and reflections on technology acceptance research". *Journal of the Association for Information Systems*, *8*(4), 204–205.

Hjort, N. L. & Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association*, *98*(464), 879–899.

Ho, M.-h. R., Stark, S., & Chernyshenko, O. (2012). Graphical representation of structural equation models using path diagrams. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 43–55). New York: Guilford Press.

Hobbs, R. (1998). The seven great debates in the media literacy movement. *Journal of Communication*, *48*(1), 16–32. doi:10.1111/j.1460-2466.1998.tb02734.x

Holsapple, C. W. & Wu, J. (2007). User acceptance of virtual worlds: The hedonic framework. *ACM SIGMIS Database*, *38*(4), 86. doi:10.1145/1314234.1314250

Hox, J. J., Maas, Cora J. M., & Brinkhuis, Matthieu J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*(2), 157–170.

Hoyle, R. H. (Ed.). (2012a). *Handbook of structural equation modeling.* New York: Guilford Press.

Hoyle, R. H. (2012b). Introduction and overview. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 3–16). New York: Guilford Press.

Hoyle, R. H. (2012c). Model specification in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 126–144). New York: Guilford Press.

Hoyle, R. H. & Isherwood, J. C. (2013). Reporting results from structural equation modeling analyses in Archives of Scientific Psychology. *Archives of Scientific Psychology*, *1*(1), 14–22. doi:10.1037/arc0000004

Hsu, C.-L. & Lin, J. C.-C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management*, *45*(1), 65–74. doi:10.1016/j.im.2007.11.001

Hsu, C.-L. & Lu, H.-P. (2004). Why do people play on-line games? An extended TAM with social influences and flow experience. *Information & Management*, *41*(7), 853–868. doi:10.1016/j.im.2003.08.014

Hu, L.-t., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351–362. doi:10.1037/0033-2909.112.2.351

Hubbard, R. (1995). The earth is highly significantly round (p<.0001). *American Psychologist*, *50*(12), 1098.

Hundley, H. L. & Shyles, L. (2010). US teenagers' perceptions and awareness of digital technology: a focus group approach. *New Media & Society*, *12*(3), 417–433. doi:10.1177/1461444809342558

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340–362.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. doi:10.1037/0033-2909.96.1.72

Hussain, Z. & Griffiths, M. D. (2009). The attitudes, feelings, and experiences of online gamers: A qualitative analysis. *CyberPsychology & Behavior*, *12*(6), 747–753. doi:10.1089/cpb.2009.0059

Hwang, M. I. (1998). Did task type matter in the use of decision room GSS? A critical review and a meta-analysis. *Omega*, *26*(1), 1–15. doi:10.1016/S0305-0483(97)00047-9

IBM AMOS Support. (2009). AMOS tests for multivariate normality with missing data. Retrieved November 13, 2013, from http://www-01.ibm.com/support/docview.wss?uid=swg21479360

IBM AMOS Support. (2010). Satorra-Bentler scaled chi-square test statistic. Retrieved February 10, 2014, from http://www-01.ibm.com/support/docview.wss?uid=swg21476184

IBM AMOS Support. (2012a). Box-Cox transformations for regression. Retrieved February 21, 2014, from http://www-01.ibm.com/support/docview.wss?uid=swg21476027

IBM AMOS Support. (2012b). SRMR is not printed when there is missing data. Retrieved February 21, 2014, from http://www-01.ibm.com/support/docview.wss?uid=swg21614980

IBM Software Group. (2013). SPSS missing values for Windows, version 22.0: Manual. Armonk, NY: IBM Corporation.

Ilie, A., Ioan, S., Zagrean, L., & Moldovan, M. (2008). Better to be red than blue in virtual competition [Rapid communication]. *CyberPsychology & Behavior*, *11*(3), 375–377. doi:10.1089/cpb.2007.0122

Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2013). Examining the psychometric properties of the Mayer-Salovey-Caruso Emotional Intelligence Test: Findings from an Eastern European culture. *European Journal of Psychological Assessment*, *29*(2), 121–128. doi:10.1027/1015-5759/a000132

In'nami, Y. & Koizumi, R. (2013). Review of sample size for structural equation models in second language testing and learning research: A Monte Carlo approach. *International Journal of Testing*, *13*(4), 329–353. doi:10.1080/15305058.2013.806925

Irons, J. L., Remington, R. W., & McLean, J. P. (2011). Not so fast: Rethinking the effects of action video games on attentional capacity. *Australian Journal of Psychology*, *63*(4), 224–231. doi:10.1111/j.1742-9536.2011.00001.x

Jaccard, J. & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*(2), 348–357. doi:10.1037/0033-2909.117.2.348

Jackson, S. A., Thomas, P. R., Marsh, H. W., & Smethurst, C. J. (2001). Relationships between flow, self-concept, psychological skills, and performance. *Journal of Applied Sport Psychology*, *13*(2), 129–153. doi:10.1080/104132001753149865

Janssen, O. & Yperen, N. W. v. (2004). Employees' goal orientations, the quality of leader-member exchange, and the outcomes of job performance and job satisfaction. *The Academy of Management Journal*, *47*(3), 368–384.

Jansz, J. & Tanis, M. (2007). Appeal of playing online first person shooter games [Rapid communication]. *CyberPsychology & Behavior*, *10*(1), 133–136. doi:10.1089/cpb.2006.9981

Jarvenpaa, S. L., Leidner, D. E., Teigland, R., & Wasko, M. M. (2007). MISQ special issue on new ventures in virtual worlds [Call for papers]. *MIS Quarterly*.

Jarvenpaa, S. L. & Staples, D. S. (2000). The use of collaborative electronic media for information sharing: An exploratory study of determinants. *The Journal of Strategic Information Systems*, *9*(2-3), 129–154. doi:10.1016/S0963-8687(00)00042-1

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2012). The negative consequences of measurement model misspecification [Research note]: A Response to Aguirre-Urreta and Marakas. *MIS Quarterly*, *36*(1), 139–146.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661. doi:10.1016/j.ijhcs.2008.04.004

Jiang, Z. & Benbasat, I. (2007). Investigating the influence of the functional mechanisms of online product presentations. *Information Systems Research*, *18*(4), 454–470.

Johnson, D. W., Johnson, R. T., & Roseth, C. J. (2012). Competition and performance: More facts, more understanding? Comment on Murayama and Elliot (2012). *Psychological Bulletin*, *138*(6), 1071–1078. doi:10.1037/a0029454

Jones, S. R. & Fernyhough, C. (2007). Reliability of factorial structure of the Peters et al. delusions inventory (PDI-21). *Personality and Individual Differences*, *43*(4), 647–656. doi:10.1016/j.paid.2007.01.018

Joseph, D. L. & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, *95*(1), 54–78. doi:10.1037/a0017286

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology*, *92*(1), 107–127. doi:10.1037/0021-9010.92.1.107

Kadavasal, M. S., Dhara, K. K., Wu, X., & Krishnaswamy, V. (2007). Mixed reality for enhancing business communications using virtual worlds. In S. N. Spencer (Ed.). New York: Association for Computing Machinery.

Kamakura, W. A. & Wedel, M. (2000). Factor analysis and missing data. *Journal of Marketing Research*, *37*(4), 490–498.

Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills & self-regulation for learning: A trait perspective. *Special Issue: A Symposium on Self-Regulated Learning*, *8*(3), 185–209. doi:10.1016/S1041-6080(96)90014-X

Kaplan, D. (1995). Statistical power in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 100–117). Thousand Oaks: Sage Publications.

Kaplan, D. & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). New York: Guilford Press.

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, *65*(8), 724–736. doi:10.1016/j.ijhcs.2007.02.003

Kato, P. M. (2010). Video games in health care: Closing the gap. *Review of General Psychology*, *14*(2), 113–121. doi:10.1037/a0019441

Katz, E. (1959). Mass communications research and the study of popular culture: An editorial note on a possible future for this journal. *Studies in Public Communication*, *2*.

Katz, E. & Foulkes, D. (1962). On the use of the mass media as "escape": Clarification of a concept. *The Public Opinion Quarterly*, *26*(3), 377–388.

Katz, E., Haas, H., & Gurevitch, M. (1973). On the Use of the Mass Media for Important Things. *American Sociological Review*, *38*(2), 164–181.

Kelly, A. B., O'Flaherty, M., Connor, J. P., Homel, R., Toumbourou, J. W., Patton, G. C., & Williams, J. (2011). The influence of parents, siblings and peers on pre- and early-teen smoking: A multilevel model. *Drug & Alcohol Review*, *30*(4), 381–387. doi:10.1111/j.1465-3362.2010.00231.x

Kelly, J. R. & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes*, *86*(1), 99–130. doi:10.1006/obhd.2001.2974

Kenny, D. A. & Milan, S. (2012). Identification: A nontechnical discussion of a technical issue. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 145–163). New York: Guilford Press.

Kiang, M. Y. & Kumar, A. (2001). An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications. *Information Systems Research*, *12*(2), 177.

Kientz, J. A., Hayes, G. R., Westeyn, T. L., Starner, T., & Abowd, G. D. (2007). Pervasive computing and autism: Assisting caregivers of children with special needs. *IEEE Pervasive Computing*, *6*, 28–35. doi:10.1109/MPRV.2007.18

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, *8*(1), 13–24. doi:10.1016/j.iheduc.2004.12.001

Kim, H.-W., Chan, H. C., & Kankanhalli, A. (2012). What motivates people to purchase digital items on virtual community websites? The desire for online self-presentation [Abstract]. *Information Systems Research*, *23*(4), 1232–1245. doi:10.1287/isre.1110.0411

Kim, H.-W. & Kankanhalli, A. (2009). Investigating user resistance to Information Systems implementation: A status quo bias perspective. *MIS Quarterly*, *33*(3).

Kim, J. K., Brick, J. M., Fuller, W. A., & Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society: Series B*, *68*(3), 509–521. doi:10.1111/j.1467-9868.2006.00546.x

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 368–390. doi:10.1207/s15328007sem1203_2

Kim, S. (2009). The integrative framework of technology use: An extension and test. *MIS Quarterly*, *33*(3), 513–537.

Kim, T.-Y., Cable, D. M., Kim, S.-P., & Wang, J. (2009). Emotional competence and work performance: The mediating effect of proactivity and the moderating effect of job autonomy. *Journal of Organizational Behavior*, *30*(7), 983–1000. doi:10.1002/job.610

Kim, T. & Biocca, F. (1997). Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication*, *3*(2). doi:10.1111/j.1083-6101.1997.tb00073.x

Kingston, E. (2008). Emotional competence and drop-out rates in higher education. *Education + Training*, *50*(2), 128–139. doi:10.1108/00400910810862119

Klimmt, C., Hefner, D., & Vorderer, P. (2009). The video game experience as "true" identification: A theory of enjoyable alterations of players' self-perception. *Communication Theory*, *19*(4), 351–373. doi:10.1111/j.1468-2885.2009.01347.x

Klimmt, C., Schmid, H., & Orthmann, J. (2009). Exploring the enjoyment of playing browser games [Rapid communication]. *CyberPsychology & Behavior*, *12*(2), 231–234. doi:10.1089/cpb.2008.0128

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kline, R. B. (2012). Assumptions in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 111–125). New York: Guilford Press.

Kohler, T., Fueller, J., Matzler, K., & Stieger, D. (2011). Co-creation in virtual worlds: The design of the user experience. *MIS Quarterly*, *35*(3), 773–788.

Korsgaard, M. A., Picot, A. O., Wigand, R. T., Welpe, I. M., & Assmann, J. J. (2010). Cooperation, coordination, and trust in virtual teams: Insights from virtual games: Chapter 20. In W. S. Bainbridge (Ed.), *Online worlds* (pp. 253–264). London: Springer.

Krueger, J. & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, *82*(2), 180–188. doi:10.1037/0022-3514.82.2.180

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*(1), 148–161. doi:10.1037/0022-3514.86.1.148

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can lottery incentives boost web survey response rates? Findings from four experiments. *Research in Higher Education*, *52*(5), 537–553. doi:10.1007/s11162-010-9203-2

Lai, K. & Kelley, K. (2011). Accuracy in parameter estimation for targeted effects in structural equation modeling: Sample size planning for narrow confidence intervals. *Psychological Methods*, *16*(2), 127–148. doi:10.1037/a0021764

Lamb, R. & Kling, R. (2003). Reconceptualizing users as social actors in Information Systems research. *MIS Quarterly*, *27*(2).

Lane, A. M., Meyer, B. B., Devonport, T. J., Davies, K. A., Thelwell, R., Gill, G. S., … Weston, N. (2009). Validity of the emotional intelligence scale for use in sport. *Journal of Sports Science and Medicine*, *8*(2), 289–295.

Lass-Hennemann, J., Deuter, C. E., Kuehl, L. K., Schulz, A., Blumenthal, T. D., & Schachinger, H. (2010). Effects of stress on human mating preferences: Stressed individuals prefer dissimilar mates. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1691), 2175–2183. doi:10.1098/rspb.2010.0258

Law, K. S., Wong, C.-S., & Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology*, *89*(3), 483–496. doi:10.1037/0021-9010.89.3.483

Lee, A. S. & Baskerville, R. L. (2003). Generalizing generalizability in Information Systems research. *Information Systems Research*, *14*(3), 221–243. doi:10.1287/isre.14.3.221.16560

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x

Lee, T., Cai, L., & MacCallum, R. C. (2012). Power analysis for tests of structural equation models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 181–194). New York: Guilford Press.

Lee, Y., Kozar, K. A., & Larsen, K. R. (2009). Avatar e-mail versus traditional e-mail: Perceptual difference and media selection difference. *Decision Support Systems*, *46*(2), 451–467. doi:10.1016/j.dss.2007.11.008

LePine, J. A. & Dyne, L. v. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with Big Five personality characteristics and cognitive ability. *Journal of Applied Psychology*, *86*(2), 326–336. doi:10.1037//0021-9010.86.2.326

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence inventory. *Presence: Teleoperators and Virtual Environments*, *10*(3), 282–297. doi:10.1162/105474601300343612

Leung, L. & Lee, P. S. (2005). Multiple determinants of life quality: The roles of Internet activities, use of new media, social support, and leisure activities. *Telematics and Informatics*, *22*(3), 161–180. doi:10.1016/j.tele.2004.04.003

Li, X. & Beretvas, S. N. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(2), 241–264. doi:10.1080/10705511.2013.769391

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 1–55.

Lim, K. H. & Benbasat, I. (2000). The effect of multimedia on perceived equivocality and perceived usefulness of Information Systems. *MIS Quarterly*, *24*(3), 449–471.

Lim, K. H., Benbasat, I., & Ward, L. M. (2000). The role of multimedia in changing first impression bias. *Information Systems Research*, *11*(2), 115–136. doi:10.1287/isre.11.2.115.11776

Lim, S. & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, *68*(1-2), 57–68. doi:10.1016/j.ijhcs.2009.09.008

Limayem, M., Hirt, S. G., & Cheung, C. (2008). How habit limits the predictive power of intention: The case of Information Systems continuance. *MIS Quarterly*, *31*(4), 705–737.

Little, R. J. A. (1988). A test of Missing Completely at Random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. doi:10.2307/2290157

Liu, H. (2010). Dynamics of pricing in the video game console market: Skimming or penetration? *Journal of Marketing Research*, *47*(3), 428–443. doi:10.1509/jmkr.47.3.428

Locke, E. A. (2005). Why emotional intelligence is an invalid concept. *Journal of Organizational Behavior*, *26*(4), 425–431. doi:10.1002/job.318

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(3), 635–694. doi:10.2466/pr0.1957.3.3.635

Lucas, H. C. J. (1975). Performance and the use of an information systems. *Management Science*, *21*(8), 908–919.

Lucas, H. C. J., Burton Swanson, E., & Zmud, R. W. (2007). Implementation, innovation, and related themes over the years in Information Systems research. *Journal of the Association for Information Systems*, *8*(4), 206–210.

Lyons, B. D., Hoffman, B. J., & Michel, J. W. (2009). Not much more than g? An examination of the impact of intelligence on NFL performance. *Human Performance*, *22*(3), 225–245. doi:10.1080/08959280902970401

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 16–36). Thousand Oaks: Sage Publications.

MacCallum, R. C. (2003). 2001 Presidential address: Working with imperfect models. *Multivariate behavioral research*, *38*(1), 113–139. doi:10.1207/S15327906MBR3801_5

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*, *11*(1), 19–35.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149.

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*(3), 340–345.

MacCallum, R. C., Lee, T., & Browne, M. W. (2010). The issue of isopower in power analysis for tests of structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(1), 23–41. doi:10.1080/10705510903438906

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99.

MacCann, C. (2010). Further examination of emotional intelligence as a standard intelligence: A latent variable analysis of fluid intelligence, crystallized intelligence, and emotional intelligence. *Personality and Individual Differences*, *49*(5), 490–496. doi:10.1016/j.paid.2010.05.010

MacDorman, K. F. (2006a). Introduction to the special issue on android science. *Connection Science*, *18*(4), 313–317. doi:10.1080/09540090600906258

MacDorman, K. F. (2006b). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In K. F. MacDorman & H. Ishiguro (Eds.), *Toward social mechanisms of android science*. Mahwah, New Jersey, USA: Lawrence Erlbaum.

Mackelprang, A. J. (1970). Missing data in factor analysis and multiple regression. *Midwest Journal of Political Science*, *14*(3), 493–505. doi:10.2307/2110316

MacKenzie, S. B. & Podsakoff, P. M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. *Journal of Retailing*, *88*(4), 542–555. doi:10.1016/j.jretai.2012.08.001

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90*(4), 710–730. doi:10.1037/0021-9010.90.4.710

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293–334.

Majchrzak, A., Beath, C. M., Lim, R., & Chin, W. W. (2005). Managing client dialogues during Information Systems design to facilitate client learning. *MIS Quarterly*, *29*(4), 653–2672.

Malone, P. S. & Lubansky, J. B. (2012). Preparing data for structural equation modeling: Doing Your Homework. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 263–276). New York: Guilford Press.

Mantymaki, M. & Riemer, K. (2011). How social are social virtual worlds? An investigation of hedonic, utilitarian, social and normative usage drivers: Paper 126. In P. B. Seddon & S. Gregor (Eds.), *Proceedings of the Fifteenth Pacific Asia Conference on Information Systems (PACIS 2011)*. Queensland University of Technology.

Marcoulides, G. A., Chin, W. W., & Saunders, C. S. (2009). A critical look at partial least squares modeling. *MIS Quarterly*, *33*(1), 171–175.

Marcoulides, G. A. & Ing, M. (2012). Automated Structural Equation Modeling Strategies. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 690–704). New York: Guilford Press.

Marcoulides, G. A. & Saunders, C. S. (2006). PLS: A silver bullet? [Editor's comments]. *MIS Quarterly*, *30*(2), 211.

Markus, K. A. (2010). Structural equations and causal explanations: Some challenges for causal SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(4), 654–676. doi:10.1080/10705511.2010.510068

Marsh, H. W. & Jackson, S. A. (1999). Flow experience in sport: Construct validation of multidimensional, hierarchical state and trait responses. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(4), 343–371. doi:10.1080/10705519909540140

Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–491. doi:10.1037/a0019227

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476. doi:10.1080/10705510903008220

Marsh, H. W., Wen, Z., Nagengast, B., & Hau, K.-T. (2012). Structural equation models of latent interaction. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 436–458). New York: Guilford Press.

Massey, A. P. & Montoya-Weiss, M. M. (2006). Unraveling the temporal fabric of knowledge conversion: A model of media selection and use. *MIS Quarterly*, *30*(1), 99–114.

Mayer, A. (2009). Online social networks in economics. *Decision Support Systems*, *47*(3), 169–184. doi:10.1016/j.dss.2009.02.009

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, *27*(4), 267–298. doi:10.1016/S0160-2896(99)00016-1

Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Selecting a measure of emotional intelligence: The case for ability scales. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 320–342). San Francisco: Jossey-Bass.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*(1), 507–536. doi:10.1146/annurev.psych.59.103006.093646

Mayer, J. D. & Salovey, P. (1997). What is emotional intelligence. In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence* (pp. 3–34). New York, NY: Basic Books.

McDonald, R. P. & Ho, M.-h. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. doi:10.1037/1082-989X.7.1.64

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. doi:10.1016/j.intell.2008.08.004

McGrew, K. S. & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 651–675. doi:10.1002/pits.20497

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York: Guilford Press.

Mennecke, B. E., McNeill, D., Ganis, M., Roche, E. M., Bray, D. A., Konsynski, B. R., ... Lester, J. (2008). Second Life and other Virtual Worlds: A Roadmap for Research: Article 20. *Communications of the Association for Information Systems (CAIS)*, *22*, 371–388.

Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K., & Perelgut, S. (2009). Virtual worlds–Past, present, and future: New directions in social computing. *Decision Support Systems*, *47*(3), 204–228. doi:10.1016/j.dss.2009.02.014

Mikels, J. A., Löckenhoff, C. E., Maglio, S. J., Carstensen, L. L., Goldstein, M. K., & Garber, A. (2010). Following your heart or your head: Focusing on emotions versus information differentially influences the decisions of younger and older adults. *Journal of Experimental Psychology: Applied*, *16*(1), 87–95. doi:10.1037/a0018500

Miles, J. (2003). A framework for power analysis using a structural equation modelling procedure. *BMC Medical Research Methodology*, *3*, 27–11. doi:10.1186/1471-2288-3-27

Millsap, R. E. & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380–392). New York: Guilford Press.

Miranda, S. M. & Saunders, C. S. (2003). The social construction of meaning: An alternative perspective on information sharing. *Information Systems Research*, *14*(1), 87–106. doi:10.1287/isre.14.1.87.14765

Mithas, S. & Krishnan, M. S. (2009). From association to causation via a potential outcomes approach. *Information Systems Research*, *20*(2), 295–313.

Moon, J. Y. & Sproull, L. S. (2008). The role of feedback in managing the Internet-based volunteer work force. *Information Systems Research*, *19*(4), 494–515. doi:10.1287/isre.1080.0208

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley. *IEEE Robotics and Automation*, *19*(2), 98–100. doi:10.1109/MRA.2012.2192811

Morris, J. H., Neuwirth, C. M., Regli, S. H., Chandhok, R., & Wenger, G. C. (1999). Interface issues in computer support for asynchronous communication. *ACM Computing Surveys*, *31*(2es), 11–es. doi:10.1145/323216.323358

Mplus Support. (2007). SEM with multiple imputation. Retrieved January 22, 2014, from http://www.statmodel.com/discussion/messages/11/2518.html?1331238843

Mplus Support. (2008). MLM (Satorra-Bentler chi-square): Nonormal and missing. Retrieved February 11, 2014, from http://www.statmodel.com/discussion/messages/11/3542.html?1220887237

Mueller, J., Hutter, K., Fueller, J., & Matzler, K. (2011). Virtual worlds as knowledge management platform - a practice-perspective. *Information Systems Journal*, *21*(6), 479–501. doi:10.1111/j.1365-2575.2010.00366.x

Murayama, K. & Elliot, A. J. (2012). The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin*, *138*(6), 1035–1070. doi:10.1037/a0028324

Muteki, K., MacGregor, J. F., & Ueda, T. (2005). Estimation of missing data using latent variable methods with auxiliary information. *Chemometrics and Intelligent Laboratory Systems*, *78*(1–2), 41–50. doi:10.1016/j.chemolab.2004.12.004

Muthén, B. O., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, *16*(1), 17–33. doi:10.1037/a0022634

Muthén, B. O. & Kaplan, D. (1985). A comparison of methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189. Retrieved from http://www.statmodel.com/download/MuthenKaplan.pdf

Muthén, B. O. & Kaplan, D. (1992). A comparison of methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19–30. Retrieved from http://www.statmodel.com/bmuthen/articles/Article_040.pdf

Muthén, L. K. & Muthén, B. O. (1998-2010). *Mplus User's Guide. Sixth Edition.* Los Angeles, CA: Muthén & Muthén.

Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures*, *5*(4), 297–310. doi:10.1080/19312458.2011.624490

Nah, F. F.-H., Eschenbrenner, B., & DeWester, D. (2011). Enhancing brand equity through flow and telepresence: A comparison of 2D and 3D virtual worlds. *MIS Quarterly*, *35*(3), 731–747.

Nakagawa, S. & Freckleton, R. P. (2008). Missing inaction: The dangers of ignoring missing data. *Trends in Ecology & Evolution*, *23*(11), 592–596. doi:10.1016/j.tree.2008.06.014

Nash, E. B., Edwards, G. W., Thompson, J. A., & Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of Human-Computer Interaction*, *12*(1), 1–41. doi:10.1207/S15327590IJHC1201_1

Nevo, S., Nevo, D., & Carmel, E. (2011). Unlocking the business potential of virtual worlds. *MIT Sloan Management Review*, *52*(3), 14–17.

Newton, J. H. & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools*, 621–634. doi:10.1002/pits.20495

Nguyen, V.-H., Merienne, F., & Martinez, J.-L. (2010). Training based on real-time motion evaluation for functional rehabilitation in virtual environment. *International Journal of Image and Graphics*, *10*(02), 235. doi:10.1142/S0219467810003755

Ngwenyama, O. K. & Lee, A. S. (1997). Communication richness in electronic mail: Critical social theory and the contextuality of meaning. *MIS Quarterly*, *21*(2), 145–167.

Niederman, F., Beise, C. M., & Beranek, P. M. (1996). Issues and concerns about computer-supported meetings: The facilitator's perspective. *MIS Quarterly*, *20*(1), 1–22.

Nieuwdorp, E. (2007). The pervasive discourse. *Computers in Entertainment*, *5*(2), 13. doi:10.1145/1279540.1279553

Nikken, P. (2003). Parental mediation of children's video game playing: A similar construct as television mediation. In *Proceedings of the 2003 Digital Games Research Association Conference (DiGRA 2003)*. Utrecht: University of Utrecht.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396–402.

Oertzen, T. v. (2010). Power equivalence in structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, *63*(2), 257–272. doi:10.1348/000711009X441021

Olsson, U. H., Foss, T., & Troye, S. V. (2003). Does the ADF fit function decrease when the kurtosis increases? *British Journal of Mathematical and Statistical Psychology*, *56*(2), 289–303.

Publication manual of the American Psychological Association (6th ed.). (2011). 5th print. Washington, DC: American Psychological Association.

Ortiz de Guinea, A. & Markus, M. L. (2009). Why break the habit of a lifetime? Rethinking the roles of intention, habit, and emotion in continuing information technology use. *MIS Quarterly*, *33*(3), 433–444.

Orton, N. J. H. & Ipsitz, S. R. L. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician*, *55*(3).

Ory, D. T. & Mokhtarian, P. L. (2010). The impact of non-normality, sample size and estimation technique on goodness-of-fit measures in structural equation modeling: Evidence from ten empirical models of travel behavior. *Quality & Quantity*, *44*(3), 427–445. doi:10.1007/s11135-008-9215-6

Oster, D., Schoder, D., Putzke, J., Fischbach, K., Gloor, P. A., & Steinfels, S. (2010). Tell your customers what they really want to hear: Improving the effectiveness of advertising campaigns in the financial sector using SNA on the Web2.0. In *Proceedings of the 2010 International Sunbelt Social Network Conference (Sunbelt XXX).*

Oster, D., Steinfels, S., Putzke, J., Fischbach, K., Gloor, P. A., & Schoder, D. (2009). Measuring and enhancing advertising success using SNA on the Web. In *Proceedings of the 2009 International Sunbelt Social Network Conference (Sunbelt XXIX).*

Overby, E., Slaughter, S. A., & Konsynski, B. (2010). The design, use, and consequences of virtual processes [Research commentary, abstract]. *Information Systems Research*, *21*(4), 700–710. doi:10.1287/isre.1100.0319

Paddock, S. M. (2002). Bayesian nonparametric multiple imputation of partially observed data with ignorable nonresponse. *Biometrika*, *89*(3), 529–538.

Palomo, J., Dunson, D. B., & Bollen, K. A. (2007). Bayesian structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (Vol. 1, pp. 163–188). Handbook of computing and statistics with applications. Amsterdam: Elsevier. Retrieved from http://books.google.de/books?id=Kxv1FLSZ_eIC

Papargyris, A. & Poulymenakou, A. (2005). Learning to fly in persistent digital worlds: The case of massively multiplayer online role playing games. *SIGGROUP Bulletin*, *25*(1), 41–49.

Parasuraman, R., Visser, E. d., Clarke, E., McGarry, W. R., Hussey, E., Shaw, T., & Thompson, J. C. (2009). Detecting threat-related intentional actions of others: Effects of image quality, response mode, and target cuing on vigilance. *Journal of Experimental Psychology: Applied*, *15*(4), 275–290. doi:10.1037/a0017132

Pargman, D. & Erissson, A. (2005). Law, order and conflicts of interest in massively multiplayer online games. In *Proceedings of the 2005 Digital Games Research Association Conference (DiGRA 2005).*

Park, N., Min Lee, K., Annie Jin, S.-A., & Kang, S. (2010). Effects of pre-game stories on feelings of presence and evaluation of computer games. *International Journal of Human-Computer Studies*, *68*(11), 822–833. doi:10.1016/j.ijhcs.2010.07.002

Partala, T. (2011). Psychological needs and virtual worlds: Case Second Life. *International Journal of Human-Computer Studies*, *69*(12), 787–800. doi:10.1016/j.ijhcs.2011.07.004

Parwoll, M. & Wagner, R. (2012). The impact of missing values on PLS model fitting [Abstract]. In W. A. Gaul, A. Geyer-Schulz, L.

Schmidt-Thieme, & J. Kunze (Eds.), *Challenges at the interface of data analysis, computer science, and optimization* (pp. 537–544). Studies in Classification, Data Analysis, and Knowledge Organization. Berlin: Springer.

Patil, V. H., Singh, S. N., Mishra, S., & Todd Donavan, D. (2008). Efficient theory development and factor retention criteria: Abandon the "eigenvalue greater than one" criterion. *Journal of Business Research*, *61*(2), 162–170. doi:10.1016/j.jbusres.2007.05.008

Paul, D. L. & McDaniel, Reuben R. , Jr. (2004). A field study of the effect of interpersonal trust on virtual collaborative relationship performance. *MIS Quarterly*, *28*(2), 183–227.

Pavlou, P. A. & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, *15*(1), 37–59. doi:10.1287/isre.1040.0015

Perugini, M. & Bagozzi, R. P. (2001). The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour. *The British journal of social psychology / the British Psychological Society*, *40*(Pt 1), 79–98.

Peter, F. (2005). Identification of multivariate outliers: A performance study. *Austrian Journal of Statistics*, *34*(2), 127–138.

Peters, E., Dieckmann, N. F., Västfjäll, D., Mertz, C. K., Slovic, P., & Hibbard, J. H. (2009). Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied*, *15*(3), 213–227. doi:10.1037/a0016978

Petrakou, A. (2010). Interacting through avatars: Virtual worlds as a context for online education. *Computers & Education*, *54*(4), 1020–1027. doi:10.1016/j.compedu.2009.10.007

Petrides, K. V. & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, *17*(1), 39–57. doi:10.1002/per.466

Petter, S., Rai, A., & Straub, D. W. (2012). The critical importance of construct measurement specification: A response to Aguirre-Urreta and Marakas [Research note]. *MIS Quarterly*, *36*(1), 147–156.

Petter, S., Straub, D. W., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623–656.

Phillips, D. L. & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology*, *77*(5), 921–940.

Phillips, J. M. & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology*, *82*(5), 792–802. doi:10.1037/0021-9010.82.5.792

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly*, *25*(4), 401–426.

Pillai, R., Williams, E. A., Lowe, K. B., & Jung, D. I. (2003). Personality, transformational leadership, trust, and the 2000 U.S. presidential vote. *The Leadership Quarterly*, *14*(2), 161–192. doi:10.1016/S1048-9843(03)00008-0

Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., & Klinger, R. L. (2013). Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures. *Journal of Applied Psychology*, *98*(1), 99–113. doi:10.1037/a0029570

Podsakoff, P. M., MacKenzie, S. B., Jeong-Yeon Lee, & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879.

Polites, G. L. & Karahanna, E. (2012). Shackled to the status quo: The inhibiting effects of incumbent system habit, switching costs, and inertia on new system acceptance. *MIS Quarterly*, *36*(1), 21–42.

Porter, C. E. (2004). A typology of virtual communities: A multidisciplinary foundation for future research. *Journal of Computer-Mediated Communication*, *10*(1), 00. doi:10.1111/j.1083-6101.2004.tb00228.x

Porter, S. R. & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education*, *44*(4), 389–407. doi:10.1023/A:1024263031800

Powell, A., Piccoli, G., & Ives, B. (2004). Virtual teams: A review of current literature and directions for future research. *ACM SIGMIS Database*, *35*(1), 6–36. doi:10.1145/968464.968467

Preacher, K. J. & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, *17*(1), 1–14. doi:10.1037/a0026804

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154–166. doi:10.1037/a0019440

Ragan, E. D. (2010). The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, *19*(6), 527–543. doi:10.1162/pres_a_00016

Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research*, *19*(4), 417–433. doi:10.1287/isre.1070.0165

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, *35*(2), 299–331. doi:10.1016/S0005-7894(04)80041-8

Raykov, T. (2012). Evaluation of latent construct correlations in the presence of missing data: A note on a latent variable modelling approach. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 19–31. doi:10.1348/000711010X498162

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, *79*(4), 518–524. doi:10.1037/0021-9010.79.4.518

Remus, W. E. (1978). Testing Bowman's managerial coefficient theory using a competitive gaming environment. *Management Science*, *24*(8), 827–835.

Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2011). Trusting humans and avatars: Behavioral and neural evidence. In D. F. Galletta & T.-P. Liang (Eds.), *Proceedings of the Thirty-Second International Conference on Information Systems (ICIS 2011)*.

Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, *36*(1), iiv–8.

Rivard, S. (2014). The ions of theory construction [Editor's comments]. *MIS Quarterly*, *38*(2), iii–xiv.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288. doi:10.1037/0033-2909.130.2.261

Roberts, N., Thatcher, J. B., & Grover, V. (2010). Advancing operations management theory using exploratory structural equation modelling techniques. *International Journal of Production Research*, *48*(15), 4329–4353. doi:10.1080/00207540902991682

Rogow, F. (2004). Shifting from media to literacy: One opinion on the challenges of media literacy education. *American Behavioral Scientist*, *48*, 30–34.

Roquilly, C. (2011). Control over virtual worlds by game companies: Issues and recommendations. *MIS Quarterly*, *35*(3), 653–671.

Rosenbloom, A. (2004). Interactive immersion in 3D computer graphics. *Communications of the ACM*, *47*(8), 28. doi:10.1145/1012037.1012058

Rosser, J., Lynch, P. J., Cuddihy, L., Gentile, D. A., Klonsky, J., & Merrell, R. (2007). The impact of video games on training surgeons in the 21st century. *Archives of Surgery*, *142*(2), 181–6, discusssion 186. doi:10.1001/archsurg.142.2.181

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*(3), 537–560. doi:10.1111/j.1744-6570.1994.tb01736.x

Roth, P. L., Switzer, F. S. I., & Switzer, D. M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods*, *2*(3), 211–232. doi:10.1177/109442819923001

Rowe, R. M. & McKenna, F. P. (2001). Skilled anticipation in real-world tasks: Measurement of attentional demands in the domain of tennis. *Journal of Experimental Psychology: Applied*, *7*(1), 60–67. doi:10.1037//1076-898X.7.1.60

Rubin, A. M. (1998). Media literacy. *Journal of Communication*, *48*(1), 3–4. doi:10.1111/j.1460-2466.1998.tb02732.x

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. doi:10.1093/biomet/63.3.581

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489.

Rudner, L. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, *12*(2), 1–11.

Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st century. *Mass Communication & Society*, *3*(1), 3–37.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. doi:10.1037/0033-295X.110.1.145

Sacau, A., Laarni, J., & Hartmann, T. (2008). Influence of individual factors on presence. *Computers in Human Behavior*, *24*(5), 2255–2273. doi:10.1016/j.chb.2007.11.001

Salen, K. & Zimmerman, E. (2004). *Rules of play: Game design fundamentals* (3rd ed.). Cambridge, Mass: MIT Press.

Salovey, P. & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, *9*(3), 185–211.

Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399–419). Thousand Oaks, CA, US: Sage Publications.

Satorra, A. & Bentler, P. M. (August 3 / 1999). A scaled difference chi-square test statistic for moment structure analysis [Working paper]. Retrieved from http://www.escholarship.org/uc/item/23c604tb

Saunders, C. S., Rutkowski, A. F., Genuchten, M. v., Vogel, D., & Orrego, J. M. (2011). Virtual space and place: Theory and test. *MIS Quarterly*, *35*(4), 1079–1098.

Savalei, V. & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 477–497. doi:10.1080/10705510903008238

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 3–15.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*(1), 19–35. doi:10.1111/1467-9574.00218

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. doi:10.1037/1082-989X.7.2.147

Schafer, J. L. & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate behavioral research*, *33*(4), 545–571. doi:10.1207/s15327906mbr3304_5

Schaik, P. v., Turnbull, T., Wersch, A. v., & Drummond, S. (2004). Presence within a mixed reality environment. *CyberPsychology & Behavior*, *7*, 540–552.

Scherer, K. R. (2008). Componential emotion theory can inform models of emotional competence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The Science of Emotional Intelligence* (pp. 101–126). Oxford University Press. doi:10.1093/acprof:oso/9780195181890.003.0004

Schiele, S., Weiss, T., & Putzke, J. (2011). On inter-reality literacy: Emotions as predictors of performance in virtual worlds [Research in progress]. In D. F. Galletta & T.-P. Liang (Eds.), *Proceedings of the Thirty-Second International Conference on Information Systems (ICIS 2011).*

Schroeder, R. (2002). *The social life of avatars: Presence and interaction in shared virtual environments.* London, New York: Springer.

Schroeder, R. (2006). Being there together and the future of connected presence. *Presence: Teleoperators and Virtual Environments*, *15*(4), 438–454. doi:10.1162/pres.15.4.438

Schroeder, R. & Axelsson, A.-S. (Eds.). (2006). *Avatars at work and play: Collaboration and interaction in shared virtual environments.* London: Springer.

Schubert, T. W. (2009). A New Conception of Spatial Presence: Once Again, with Feeling. *Communication Theory*, *19*(2), 161–187. doi:10.1111/j.1468-2885.2009.01340.x

Schultze, U. (2010). Embodiment and presence in virtual worlds: A review. *Journal of Information Technology*, *25*(4), 434–449. doi:10.1057/jit.2010.25

Schultze, U. (2011). The avatar as sociomaterial entanglement: A performative perspective on identity, agency and world-making in virtual worlds. In D. F. Galletta & T.-P. Liang (Eds.), *Proceedings of the Thirty-Second International Conference on Information Systems (ICIS 2011).*

Schultze, U. & Leahy, M. M. (2009). The avatar-self relationship: Enacting presence in Second Life. In *Proceedings of the Thirtieth International Conference on Information Systems (ICIS 2009).* Retrieved from http://aisel.aisnet.org/icis2009/12

Schultze, U. & Orlikowski, W. J. (2010). Virtual worlds: A performative perspective on globally distributed, immersive work [Research commentary]: Research Commentary. *Information Systems Research*, *21*(4), 810–821. doi:10.1287/isre.1100.0321

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*(2), 167–177. doi:10.1016/S0191-8869(98)00001-4

Schwartz, C. M. (2011). A review of "Electronic emotion: The mediation of emotion via information and communication technologies". *The Information Society*, *27*(4), 279–280. doi:10.1080/01972243.2011.583849

Schwarz, A. & Chin, W. (2007). Looking forward: Toward an understanding of the nature and definition of IT acceptance. *Journal of the Association for Information Systems*, *8*(4), 232–243.

Schwarz, G. (2005). Overview: What is media literacy, who cares, and why? *Yearbook of the National Society for the Study of Education*, *104*(1), 5–17. doi:10.1111/j.1744-7984.2005.00002.x

Segars, A. H. & Grover, V. (1993). Re-examining perceived ease of use and usefulness: A confirmatory factor analysis. *MIS Quarterly*, *17*(4), 517–525.

Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. (2006). Video game uses and gratifications as predictors of use and game preference. In P. Vorderer & J. Bryant (Eds.), *Playing Computer Games: Motives, Responses, and Consequences* (pp. 213–224). Mahwah, NJ: Lawrence Erlbaum.

Sia, C.-L., Tan, B. C. Y., & Wei, K.-K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research*, *13*(1), 70–90.

Silva, L. (2007). Post-positivist review of Technology Acceptance Model. *Journal of the Association for Information Systems*, *8*(4), 256–266.

Silva, L. & Hirshheim, R. (2007). Fighting against windmills: Strategic Information Systems and organizational deep structures. *31*(2), 27–354.

Simon, H. A. (1972). Theories of bounded rationality. In C. B. McGuire & R. Radner (Eds.), *Decision and Organization* (pp. 161–176). North-Holland Publishing Company.

Simon, S. J., Grover, V., Teng, J. T. C., & Whitcomb, K. (1996). The relationship of Information System training methods and cognitive ability to end-user satisfaction, comprehension, and skill transfer: A longitudinal field study. *Information Systems Research*, *7*(4), 466–490. doi:10.1287/isre.7.4.466

Sims, K. (1994). Evolving virtual creatures. In *Proceedings of the Twenty-First Annual ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1994)* (pp. 15–22). Association for Computing Machinery.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*(4), 317–329. doi:10.1037/1082-989X.6.4.317

Spence, I. & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92–104. doi:10.1037/a0019491

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, Angelika. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. doi:10.1111/1467-9868.00353

Srite, M. & Karahanna, E. (2006). The role of espoused national cultural values in technology acceptance. *MIS Quarterly*, *30*(3), 679–704.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, *7*(4), 327–351. doi:10.1162/105474698565767

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. doi:10.1037/0033-2909.133.1.65

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*, *338*. doi:10.1136/bmj.b2393

Stoeber, J., Stoll, O., Pescheck, E., & Otto, K. (2008). Perfectionism and achievement goals in athletes: Relations with approach and avoidance orientations in mastery and performance goals. *Psychology of Sport and Exercise*, *9*(2), 102–121. doi:10.1016/j.psychsport.2007.02.002

Storbeck, J. (2012). Performance costs when emotion tunes inappropriate cognitive abilities: Implications for mental resources and behavior [Brief report]. *Journal of Experimental Psychology: General*, *141*(3), 411–416. doi:10.1037/a0026322

Stormshak, E. A., Bullock, B. M., Falkenstein, C. A., Kramer, L., Conger, K. J., & Falkenstein, C. A. (2009). Harnessing the power of sibling relationships as a tool for optimizing social-emotional development. In L. Kramer & K. J. Conger (Eds.), *Special Issue: Siblings as Agents of Socialization* (pp. 61–77). Wiley. doi:10.1002/cd.257

Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, *13*(2), 147–169.

Straub, D. W. & Burton-Jones, A. (2007). Veni, vidi, vici: Breaking the TAM logjam. *Journal of the Association for Information Systems*, *8*(4), 224–229.

Straub, D. W. & Limayem, M. (1995). Measuring system usage: Implications for IS theory testing. *Management Science*, *41*(8), 1328–1343.

Suddaby, R. (2011). Construct clarity in theories of management and organization [Editor's comments]. *Academy of Management Review*, *36*(3), 346–357.

Suh, K.-S., Kim, H., & Suh, E. K. (2011). What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIS Quarterly*, *35*(3), 711–729.

Surtees, A. D. R., Noordzij, M. L., & Apperly, I. A. (2012). Sometimes losing your self in space: Children's and adults' spontaneous use of multiple spatial reference frames [Brief report]. *Developmental Psychology*, *48*(1), 185–191. doi:10.1037/a0025863

Sussman, S. W. & Sproull, L. S. (1999). Straight talk: Delivering bad news through electronic communication. *Information Systems Research*, *10*(2), 150–166. doi:10.1287/isre.10.2.150

Sykes, T. A., Venkatesh, V., & Gosain, S. (2009). Model of acceptance with peer support: A social network perspective to understand employees' system use. *MIS Quarterly*, *33*(2), 371–393.

Symes, E., Ellis, R., & Tucker, M. (2005). Dissociating object-based and space-based affordances. *Visual Cognition*, *12*(7), 1337–1361. doi:10.1080/13506280444000445

Tellegen, A. & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*(3), 268–277. doi:10.1037/h0036681

Tett, R. P., Fox, K. E., & Wang, A. (2005). Development and validation of a self-report measure of emotional intelligence as a multidimensional trait domain. *Personality and Social Psychology Bulletin*, *31*(7), 859–888. doi:10.1177/0146167204272860

Thoman, E. (1999). Skills and strategies for media education. *Educational Leadership*, *56*(5), 50–54.

Tickle, J. J., Hull, J. G., Sargent, J. D., Dalton, M. A., & Heatherton, T. F. (2006). A structural equation model of social influences and exposure to media smoking on adolescent smoking. *Basic and Applied Social Psychology*, *28*(2), 117–129. doi:10.1207/s15324834basp2802_2

Timmerman, M. E. & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220.

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, *27*(2), 741–749. doi:10.1016/j.chb.2010.10.018

Truex, D. P., Baskerville, R., & Klein, H. (1999). Growing systems in emergent organizations. *Communications of the ACM*, *42*(8), 117–123. doi:10.1145/310930.310984

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management*, *24*(1), 53–62. doi:10.1016/j.jom.2005.03.001

Vakola, M., Tsaousis, I., & Nikolaou, I. (2004). The role of emotional intelligence and personality variables on attitudes toward organisational change. *Journal of Managerial Psychology*, *19*(2), 88–110. doi:10.1108/02683940410526082

Valcke, M., Bonte, S., Wever, B. d., & Rots, I. (2010). Internet parenting styles and the impact on Internet use of primary school children. *Computers & Education*, *55*(2), 454–464. doi:10.1016/j.compedu.2010.02.009

Valkenburg, P. M., Krcmar, M., Peeters, A. L., & Marseille, N. M. (1999). Developing a scale to assess three styles of television mediation: "Instructive mediation", "restrictive mediation", and "social coviewing". *Journal of Broadcasting & Electronic Media*, *43*(1), 52–66. doi:10.1080/08838159909364474

van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Medicine*, *2*(10), e267. doi:10.1371/journal.pmed.0020267

van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation for item scores when test data are factorially complex. *British Journal of Mathematical and Statistical Psychology*, *60*(2), 315–337. doi:10.1348/000711006X117574

Vasilakos, A. V., Wei, L., Nguyen, T. H. D., Thien Qui, T. C., Chen, L. C., Boj, C., … Marentakis, G. (2008). Interactive theatre via mixed reality and Ambient Intelligence. *Information Sciences*, *178*(3), 679–693. doi:10.1016/j.ins.2007.08.029

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the Technology Acceptance Model. *Information Systems Research*, *11*(4), 342–365.

Venkatesh, V., Davis, F. D., & Morris, M. G. (2007). Dead or alive? The development, trajectory and future of technology adoption research. *Journal of the Association for Information Systems*, *8*(4), 268–286.

Venkatesh, V., L. Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478.

Verhagen, T., Feldberg, F., Hooff, B. v. d., & Meents, S. (2009). Understanding virtual world usage: A multipurpose model and empirical testing. In *Proceedings of the Seventeenth European Conference on Information Systems (ECIS 2009)*.

Vermunt, J. K., van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociological Methodology*, *38*(1), 369–397. doi:10.1111/j.1467-9531.2008.00202.x

Vincent, J. & Fortunati, L. (2009). *Electronic emotion: The mediation of emotion via information and communication technologies*. Interdisciplinary communication studies. Oxford, New York: Peter Lang. Retrieved from http://books.google.ie/books?id=ZgFSeV4oRmAC

Vinzi, V. E., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares*. Berlin: Springer. doi:10.1007/978-3-540-32827-8

Wagner, C., Schill, M., & Männer, R. (2002). Intraocular surgery on a virtual eye. *Communications of the ACM*, *45*(7). doi:10.1145/514236.514262

Wakefield, R. L., Leidner, D. E., & Garrison, G. (2008). A model of conflict, leadership, and performance in virtual teams [Research note]. *Information Systems Research*, *19*(4), 434–455. doi:10.1287/isre.1070.0149

Wakefield, R. L. & Whitten, D. (2006). Mobile computing: A user study on hedonic/utilitarian mobile device usage. *European Journal of Information Systems*, *15*(3), 292–300. doi:10.1057/palgrave.ejis.3000619

Wasko, M. M. & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, *29*(1), 35–57.

Wasko, M. M., Teigland, R., Leidner, D. E., & Jarvenpaa, S. L. (2011). Stepping into the Internet: New ventures in virtual worlds [Foreword special issue]. *MIS Quarterly*, *35*(3), 645–652.

Watson-Manheim, M. B. & Bélanger, F. (2007). Communication media repertoires: Dealing with the multiplicity of media choices. *MIS Quarterly*, *31*(2), 267–293.

Watts Sussman, S. & Schneier Siegal, W. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. *Information Systems Research*, *14*(1), 47–65.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775–808. doi:10.1037/a0027600

Webster, J. & Ho, H. (1997). Audience engagement in multimedia presentations. *ACM SIGMIS Database*, *28*(2), 63–77. doi:10.1145/264701.264706

Webster, J. & Martocchio, J. J. (1992). Microcomputer playfulness: Development of a measure with workplace implications. *MIS Quarterly*, *16*(2), 201–226.

Weiber, R. & Mühlhaus, D. (2010). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Berlin: Springer.

Weiss, T. & Schiele, S. (2013). Virtual worlds in competitive contexts: Analyzing eSports consumer needs. *Electronic Markets*, *23*(4), 307–316. doi:10.1007/s12525-013-0127-5

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 56–75). Thousand Oaks: Sage Publications.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: Guilford Press.

Wetzels, M., Odekerken-Schröder, G., & van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, *33*(1), 177–195.

Widaman, K. F. & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*(1), 16–37. doi:10.1037/1082-989X.8.1.16

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, *53*(3), 300–314. doi:10.1037/0003-066X.53.3.300

Wilk, M. B. & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, *55*(1), 1–17. doi:10.2307/2334448

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied*, *8*(4), 259–270. doi:10.1037/1076-898X.8.4.259

Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication*, *13*(4), 993–1018. doi:10.1111/j.1083-6101.2008.00428.x

Wong, C.-S., Law, K. S., & Huang, G.-h. (2008). On the importance of conducting construct-level analysis for multidimensional constructs in theory development and testing. *Journal of Management*, *34*(4), 744–764. doi:10.1177/0149206307312506

Wong, C.-S. & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude. *The Leadership Quarterly*, *13*(3), 243–274. doi:10.1016/S1048-9843(02)00099-1

Wood, R. T. A., Griffiths, M. D., & Parke, A. (2007). Experiences of time loss among videogame players: An empirical study. *CyberPsychology & Behavior*, *10*(1), 38–44. doi:10.1089/cpb.2006.9994

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Newbury Park: Sage Publications. Retrieved from http://books.google.de/books?id=FvIxxeYDLx4C

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(1), 3–24. doi:10.1080/10705519609540026

Wu, J., Li, P., & Rao, S. (2008). Why they enjoy virtual game worlds? *Journal of Electronic Commerce Research*, *9*(3), 219–230.

Xiao, B. & Benbasat, I. (2011). Product-related deception in e-commerce: A theoretical perspective. *MIS Quarterly*, *35*(1), 169–195.

Yan, Z. (2006). What Influences children's and adolescents' understanding of the complexity of the Internet? *Developmental Psychology*, *42*(3), 418–428. doi:10.1037/0012-1649.42.3.418

Yang, S.-B., Lim, J.-H., Oh, W., Animesh, A., & Pinsonneault, A. (2012). Using real options to investigate the market value of virtual world businesses [Research note, abstract]. *Information Systems Research*, *23*(3-part-2), 1011–1029. doi:10.1287/isre.1110.0397

Yee, N. (2006a). Motivations for play in online games [Rapid communication]. *CyberPsychology & Behavior*, *9*(6), 772–775. doi:10.1089/cpb.2006.9.772

Yee, N. (2006b). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, *15*(3), 309–329. doi:10.1162/pres.15.3.309

Yee, N. (2006c). The psychology of MMORPGs: Emotional investment, motivations, relationship formation, and problematic usage. In R. Schroeder & A.-S. Axelsson (Eds.), *Avatars at work and play* (Vol. 34, pp. 187–207). London: Springer.

Yi, M. Y. & Davis, F. D. (2003). Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research*, *14*(2), 146–169. doi:10.1287/isre.14.2.146.16016

Yoo, Y. & Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS Quarterly*, *25*(3), 371–390.

Young, K. (2013). A gamer's virtual world is key to understanding real-life addiction. Retrieved from http://www.goodtherapy.org/

blog/gamers-virtual-world-is-key-to-understanding-real-life-addiction-1023135

Yperen, N. W. v. & Janssen, O. (2002). Fatigued and dissatisfied or fatigued but satisfied? Goal orientations and responses to high job demands. *The Academy of Management Journal*, *45*(6), 1161–1171.

Yuan, K.-H. & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, (51), 289–309.

Yuan, K.-H. & Bentler, P. M. (1999). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions. *Statistica Sinica*, *9*, 831–853.

Yuan, K.-H. & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165–200.

Yuan, K.-H. & Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, *54*(1), 161.

Yuan, K.-H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis: The univariate case and its multivariate implication. *Sociological Methods & Research*, *34*(2), 240–258. doi:10.1177/0049124105280200

Yuan, K.-H., Lambert, P. L., & Fouladi, R. T. (2004). Mardia's multivariate kurtosis with missing data. *Multivariate behavioral research*, *39*(3), 413–437. doi:10.1207/S15327906MBR3903_2

Yung, Y. F. & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *The British journal of mathematical and statistical psychology*, *47 (Pt 1)*, 63–84.

Zhang, P. (2005). Multiple imputation of missing data with ante-dependence covariance structure. *Journal of Applied Statistics*, *32*(2), 141–155. doi:10.1080/02664760500054178

Zhao, Q. (2011). 10 scientific problems in virtual reality. *Communications of the ACM*, *54*(2), 116. doi:10.1145/1897816.1897847

Zhou, Z., Jin, X.-L., & Fang, Y. (2014). Moderating role of gender in the relationships between perceived benefits and satisfaction in social virtual world continuance. *Decision Support Systems*. Advance online publication. doi:10.1016/j.dss.2014.05.004

Zhou, Z., Jin, X.-L., Vogel, D., Guo, X., & Chen, X. (2010). Individual motivations for using social virtual worlds: An exploratory investigation in Second Life. In R. H. Sprague (Ed.), *Proceedings of the Forty-Third Annual Hawaii International Conference on System Sciences (HICSS 2010)* (pp. 1–10). New York: IEEE Press Books. doi:10.1109/HICSS.2010.230

Zhu, L., Benbasat, I., & Jiang, Z. (2010). Let's shop online together: An empirical investigation of collaborative online shopping support.

Information Systems Research, *21*(4), 872–891. doi:10.1287/isre.1080.0218

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. doi:10.1037/0033-2909.99.3.432

VERZEICHNIS DER HILFSMITTEL

LEBENSLAUF

HILFSMITTEL

Zur Erstellung dieser Arbeit wurden die untenstehenden Hilfsmittel verwendet (Versionsnummern bezeichnen die jeweils letzte verwendete Version):

A. Programmieren, Parsen und Testen:
   - eclipse Indigo (3.7.2)
   - Java SE Development Kit 7 (1.7.0)
   - TortoiseSVN (1.7.5)
   - Putty (0.62)
   - Jericho HTML Parser (3.2)
   - JUnit4 (4.10)

B. Datenanalyse:
   - IBM SPSS Statistics 22
   - IBM SPSS Amos 22
   - Microsoft Excel 2010
   - R (3.0.2)

C. Typographie und Literaturverzeichnis:
   - TeXnicCenter (2.02)
   - TeXstudio (2.6.6)
   - MiKTeX (2.9)
   - Classicthesis (4.1), a LaTeX template bundle
   - Citavi (4.4.0)
   - American Psychological Association. (2011). Publication manual of the American Psychological Association (6th ed., 5th printing). Washington, DC: American Psychological Association.

LEBENSLAUF

ANGABEN ZUR PERSON

Titel, Name	Dipl.-Inform. Sabrina Schiele (geb. Steinfels)
Geburtsdatum	07. Oktober 1979
Geburtsort	Duisburg
Nationalitäten	deutsch, französisch

BERUFLICHER WERDEGANG

Apr. 08 - Okt. 12	Wissenschaftliche Mitarbeiterin und Doktorandin am Seminar für Wirtschaftsinformatik und Informationsmanagement (Prof. Dr. D. Schoder) der Universität zu Köln
Okt. 01 - Jan. 07	Studentische Hilfskraft am Lehrstuhl Informatik V, Datenbanken und Informationssysteme (Prof. Dr. M. Jarke) der Rheinisch-Westfälischen Technischen Hochschule (RWTH) Aachen

BILDUNGSWEG

Okt. 01 - Jan. 08	Diplom Informatik, RWTH Aachen (Note: 1,4) Titel der Diplomarbeit: “Innovationsmanagement und Web 2.0”
Okt. 99 - Apr. 02	Vordiplom Betriebswirtschaftslehre, Westfälische Wilhelms-Universität (WWU) Münster
Aug. 90 - Jun. 99	Abitur, Gymnasium Adolfinum, Moers (Note: 1,0)