Measuring Decision-Analytical Competence - A Psychometric Online Performance Test

Nadine Oeser

A thesis submitted to the Department of Management of the London School of Economics and Political Science for the degree of Doctor of Philosophy

London, 30th of June 2016

DECLARATION

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 85,623 words.

I can confirm that my thesis was copy edited for conventions of language, spelling and grammar by Stephen Hart (stevehartonline.com).

In support of the readability of this Ph.D. research, the author sets the continuous mentioning of the male and female description aside. Of course, each argument refers to women and men/boys and girls as decision-makers.

Nadine Oeser London, 30th of June 2016

ACKNOWLEDGEMENT

Undertaking Ph.D. research is often compared to a journey - a journey into the unknown. Indeed, the process of the present doctoral thesis meets various criteria of this analogy. Even though the final destination seemed to be clear at the beginning, several iteration loops were performed to eventually choose the topic of measuring decision-analytical competence and the way of getting there. Despite diverse time schedules, the deadline for arrival had to be adjusted as unexpected turbulences appeared. By taking some detours through at the time unfamiliar fields, the packed tools and resources had to be broadened and new skills had to be acquired to overcome obstacles. While some companions stayed during the period of travel, some left the tour, and new companions joined the journey. At the end of this bumpy and exciting trip I am happy, a little bit proud, and very thankful.

As this research project would not have happened and the corresponding thesis would not have been written without the support of a lot of smart, crucial, and reliable people, I thank everybody who facilitated and discussed my ideas, shared insights, asked critical questions, reassured me, and provided intellect, time, or contact.

For their rigorous advice, professional expertise, and operational support during my Ph.D., I thank my two supervisors Dr Gilberto Montibeller from the London School of Economics and Political Science and Loughborough University and Dr Barbara Fasolo from the London School of Economics and Political Science.

For his tremendous commitment and academic support in terms of backup in the phase of data gathering and feedback, I would like to acknowledge Dr Johannes Siebert from the University of Bayreuth. Also, I have to thank Dr Emma Soane from the London School of Economics and Political Science for her very valuable, substantial, and perceptive feedback, as well as Katharina Schmidt from Free University of Berlin for our revealing and intensive discussions about developing a psychometric test instrument. Additionally, I thank Dr Christian Wernz from Virginia Tech for his enormous help with the acquisition of test participants, which seemed to be a matter of course for him. My special thanks go to Dr James E. Matheson from Stanford University and his wonderful wife Susan, who I met at the very beginning of my research career and who turned out to be not only wise and authentic companions for my research and academic interests, but also became extended family.

One person who has spent far more than one year of his free time for my research is my former colleague and good friend Gunther Vilzmann, who programmed the complete online DAC-test. A highly considerable impact, on how this research evolved and how the decisions at some forks turned out, was made by Fay Hartrey and Maxi Becker, who not only enriched my research considerations and supported the statistical analyses, but also encouraged me and became my good friends. For his expertise in the area of design and layout I received precious help from my enduring friend Ingo Madel. All these great and unique people I would like to thank for their inspiration, impetus and joint time.

For their emotional support, reliability, and forbearance I have to thank my mom, my dad, my grandparents, and my sister. In the same breath, I would like to give props to my best friends Adrian, Florian, Jens, and Joachim for just being how they are and supporting me whenever they could.

Last but not least, I would like to say thank you to Elia and Leander for their curiosity and interest, leniency, and support and to Sarah, whom I trained in endurance and patience: for her love, understanding, encouragement, trust, and advice.

ABSTRACT

Decision-making is generally considered a key competence within organisations and for individuals. It is crucial in our daily routine and at forks in life. The latter type of situation can tremendously impact peoples' health, wealth, or happiness. Thus, the extent to which an individual is able to make sound decisions is of huge interest. Measuring this ability would enable people to assess their Decision-Making Competence (DMC) and identify areas for improvement.

Most advanced research on individual differences in DMC defines the construct mainly in terms of an individual's ability to resist decision biases - systematic deviations from normative decision rules and a concept that is mainly derived from behavioural decision theory. This research stream does not typically cover the main steps of a decision-analytical process, such as the ability to envision one's objectives, to frame a decision, or to compare alternatives. As a sound decision-making process must cover several dimensions, including not only the ability to deal with decision biases but also the ability to apply decision-analytical rules, the decision-analytical side of this construct deserves intensified investigation.

This research therefore developed a psychometric test that allows the measurement of an individual's performance by a set of six decision-analytical dimensions of DMC. On the basis of the corresponding decision-analytical literature, cognitive dimensions of analytical DMC were identified and operationalized using a catalogue of appropriate decision tasks. In two online studies with approximately 500 participants, a psychometrically sound performance test was constructed and validated.

Participants showed reasonable consistent performances across the set of Decision-Analytical Competence (DAC) tasks. An exploratory factor analysis suggested one factor underlying the presented decision tasks. A confirmatory factor analysis demonstrated acceptable model fit indices for the one-factor structure of DAC. The aggregated overall test score presented significant relationships with measures of decision-making style, fluid intelligence, and problem-solving competence.

CONTENTS

ABBREVIATIONS

Abbreviation/symbol	Explanation
English character set	
Abbr.	Abbreviation
A-DMC	Adult Decision-Making Competence index score
ALTERNATIVES	Ability to identify relevant alternatives (used in running text)
ANOVA	Analysis of variance
AOMT	Selected for the analyses of the main testing (used in tables)
AV	Average
B	Buck-passing
CFA	Confirmatory factor analysis
CFI	Comparative Fit Index
CIRMSEA	Confidence interval of RMSEA
DAC	Decision-Analytical Competence
DF	Ability to assess decision fitness (used in figures and tables)
df	Degrees of freedom
DMC	Decision-Making Competence
DO	Ability to recognise decision opportunities (used in figures and tables)
DOI	Decision Outcomes Inventory
DSE	Decision self-esteem
DU	Ability to deal with uncertainty (used in figures and tables)
EDU case	DAC-test case on *what to do after having completed school*
EFA	Exploratory factor analysis
EO	Ability to envision one's objectives (used in figures and tables)
FD	Ability to frame a decision (used in figures and tables)
FITNESS	Ability to assess decision fitness (used in running text)
FRAME	Ability to frame a decision (used in running text)
GPA	Grade point average
H	Hyper-vigilance
HMT-S	Hagen Matrices Test – short version
H_1, H_2, H_3, H_4	Research hypotheses 1-4
HU	Humboldt University of Berlin
IA	Ability to identify relevant alternatives (used in figures and tables)
II	Ability to integrate information (used in figures and tables)
IMPLEMENTATION	Ability to plan to implement a decision (used in running text)
INFORMATION	Ability to integrate information (used in running text)
INVEST case	DAC-test case on *which vehicle to buy*
IQR	Interquartile range

IoU	Intolerance of Uncertainty Scale
k	Number of answer options
LSE	London School of Economics and Political Science
M	Mean
Max	Maximum
MDMQ-I	Melbourne Decision Making Questionnaire I (Decision self-esteem)
MDMQ-II	Melbourne Decision Making Questionnaire II (Decision-making style)
Mdn	Median
Min	Minimum
Misc.	Miscellaneous
MLR	Maximum likelihood estimator with robust standard error
MT	Selected for the main testing (used in figures & tables)
N	Total number of cases
NA	Missing data/not available
n_c	Number of correct answers
n_i	Number of incorrect answers
OBJECTIVES	Ability to envision one's objectives (used in running text)
OPPORTUNITIES	Ability to recognise decision opportunities (used in running text)
O_1, O_2, O_3, O_4	Research objectives 1-4
P	Procrastination
PA-M	Point average in maths
PI	Ability to plan to implement a decision (used in figures and tables)
P_i	Item difficulty index
p_i	Item difficulty index (P_i)/100
p_{i_c}	Corrected item difficulty index (P_i)/100
Q_1, Q_2, Q_3, Q_4	Research questions 1-4
r_{it}	Discriminatory power coefficient
r_p	Pearson correlation
r_S	Spearman's rho
r_{tt}	Re-test reliability
RMSEA	Root Mean Square Error of Approximation
SC	Sub-scale
SD	Standard deviation
SE	Standard error
Skew	Skewness
SRMR	Standardized Root Mean Square Residual
SwD	Satisfaction with Decision Scale
ToH	Tower of Hanoi (task)
t	Independent samples *t*-test
UNCERTAINTY	Ability to deal with uncertainty (used in running text)

V	Vigilance
WRMR	Weighted root-mean-square residual
Y-DMC	Youth Decision-Making Competence Index
z-scores	Standardised scores with a M = 0 and SD = 1
Greek character set	
α	Cronbach's alpha
λ	Factor loadings (lambda)
λ_{stand}	Standardised factor loadings
$\lambda_{unstand}$	Unstandardised factor loadings
ω_H	McDonald's omega
X^2	Chi-square

INDEX

I

O

P

Q

R

T

V

LIST OF FIGURES

LIST OF TABLES

LIST OF FORMULAS

1 INTRODUCTION

Every day people face thousands of decisions (Keeney, 2004). Many of those decisions are fast and frugal ones (Todd & Gigerenzer, 2000), such as what to have for dinner, or which newspaper article to read. Obviously, not all decisions that people have to address would be considered "no-brainers" (Hammond, Keeney, & Raiffa, 1999, p. 2). The majority of important decisions, such as career or health choices, are demanding ones, and defined by "high stakes and serious consequences" (Hammond et al., 1999, p. 2). In such cases decision-makers' resources are finite and the decision environment appears complex, uncertain, dynamic, and competitive (Howard, 1973).

It is the latter type of decision that frequently impacts humans' lives significantly (Matheson & Howard, 1968). For this reason, it is essential to consider those kinds of decisions in a future-oriented and consequences-considering way. The often-occurring challenge for people to analytically follow through with those choices or the inconsistency in making judicious choices is present in our every-day life. The following three studies[1][2] illustrate that thoughtless choices can have devastating consequences, especially in the early careers and private lives of the young: on average more than a quarter of Germany's bachelor students cancel their studies - in mathematics it is actually every other student (Heublein et al., 2015) – 80% of German smokers start smoking before the age of 18 (Lampert, 2008) and every seventh injury or fatality in a traffic accident in Germany is caused by an individual between 18 and 24 years due to speeding (Statistisches Bundesamt, 2013).

From a long-term perspective it is not only the individual that has to deal with the consequences of such deficient choices. As demonstrated by these three examples, society is impaired by deficient individual decisions in several ways[3]: The productivity of the domestic industry is affected by the shortage of young, and particularly skilled, manpower due to breaking off of education. Indirect costs arise from health insurance for medical treatments and medication for treating the outcome of long-term smoking. And, the state is challenged by direct costs caused, for instance, by police or ambulances use rushing to car accidents.

---

[1] For the present three examples it is implied that neither quitting one's study, starting to smoke, or being involved in a car accident are desirable developments. The author is aware that there may be individual cases, in which these choices might be according to someone's personal objectives.

[2] The three examples are selected to underline the potential impact choices can have on especially young lives. Thereby the studies represent different types of decisions. In section 2.1 the specific decision type – here represented by quitting one's study - addressed by this PhD research is defined.

[3] It is important to note that there are also competent individual choices in accordance with decision makers' objectives, which may conflict with societal objectives.

Good decision-making competence (henceforth: DMC) and thereby vigilant choices could help people better achieve their long-term and well-considered objectives, thus avoid unwise decisions and reduce costs for society.

By contemplating the importance sound individual decision-making has, not only for the individual but also for society, research on DMC in general seems to be indispensable in raising public attention to the topic and to induce a greater awareness about the relevance of good individual decision-making.

Increased sensitisation in public would pave the way for and reinforce the dissemination of decision-analytical practice and therefore enable manifold applications and implementations, for instance, in education and economics. This in turn would improve individual decision-making and create better decision-makers (cf. section 2.1), which could positively affect an individual's life and, within a wider scope, our society.

Although "[n]o decision process can guarantee a perfect outcome" (Larrick, 2009, p. 461), experts from the field of decision science agree that a good decision-making process contributes to making sound decisions and achieving desired decision outcomes in the long run (e.g. Keeney, 2008; Keren & Bruine de Bruin, 2003; Larrick, Nisbett, & Morgan, 1993); accordingly, "a good decision does increase the odds of success and at the same time satisfies our very human desire to control the forces that affect our lives." (Hammond et al., 1999, p. 13) Despite the fact that decision science provides several approaches to support people in making good decisions (cf. section 2.5), *one* generally accepted definition of DMC and consequently, *one* standardised method of its assessment are missing.

The majority of prevalent research approaches the construct of DMC from behavioural decision theory. In this framework, DMC is defined and assessed primarily in terms of the ability to resist decision biases, a systematic deviation from normative decision rules. According to the literature (e.g. Baron, 2008), a good decision-making process consists of several dimensions, including not only the ability to deal with decision biases but also the ability to apply decision-analytical rules. Decision-analytical dimensions, in the form of prescriptions, such as the *ability to envision one's objectives* or the *ability to integrate information,* are typically not covered by this avenue of research. This unilateral attention to the construct of DMC – focusing on behavioural decision theory – creates a gap in prevalent research that should be addressed.

Therefore, the present research aims to approach DMC from a prescriptive perspective and to concentrate on the development of a measurement instrument capturing individual analytical decision-making. It focuses on a theory-driven definition of decision-analytical DMC that enables measuring an individual's performance on various decision-analytical dimensions by a psychometric[4] test. The resultant test instrument could allow further investigations of analytical DMC and its influence on human lives.

The following sections of this introductory chapter outline the intention of the present Ph.D. research. Section 1.1 provides a short summary of the current status of research on DMC, its measurement and why testing analytical DMC is of interest. Section 1.2 defines the research questions and presents the research objectives. Section 1.3 shows how the thesis is structured.

1.1 THE RELEVANCE OF DECISION-ANALYTICAL COMPETENCE

Decision scientists have examined decision-making behaviour and processes for decades in order to capture human decision-making processes and help people to make appropriate decisions (e.g. Milkman, Chugh, & Bazerman, 2009). Three disciplines and therefore research foci (cf. Figure 1-1) have emerged from these research approaches: the descriptive, the normative, and the prescriptive approach.

While the descriptive approach originates from behavioural psychology and examines human behaviour in terms of how humans typically make decisions (Gilovich, Griffin & Kahneman, 2002; Takemura, 2014), the normative approach stems from decision theory and focuses on the question of how decisions could be made optimally (e.g. Edwards, Miles Jr., & von Winterfeldt, 2007). The latter focuses on rational choice and normative models, which are built on basic assumptions or axioms that people should consider and that provide logical guidance for their decisions (D. E. Bell, Raiffa, & Tversky, 1988). The prescriptive approach (cf. Figure 1–1), also called decision analysis, links both the descriptive and normative approach of decision science, yet centres on the query of how people can make better decisions (e.g. D. E. Bell et al., 1988; Howard & Abbas, 2015). It focuses on helping people make better decisions by using normative models, being aware of the limitations of human judgement, and the practical problems of implementing a rational model in complex interrelations. Thus, decision analysis is understood as normative in theory and thoroughly prescriptive in practice. The overall purpose of this science is to support and enable decision-makers in business and in personal situations to make better decisions. It is settled on the link between mathematics, economics, behavioural psychology, and computer science. The roots go back to the late 1950s, first

[4] Psychometrics describes "the branch of psychology concerned with the quantification and measurement of human attributes, behavior, performance, and the like, as well as with the design, analysis, and improvement of the tests, questionnaires, and so on used in such measurement." (Zedec, 2014, p. 279)

appearing in Schlaifer's book "Probability and Statistics for Business Decisions" (1959). In the 1960s Ronald A. Howard brought decision theory and systems modelling together and gave the new discipline its name, *decision analysis* (c.f. Howard, 1968). In 1968, Howard Raiffa published the first book bearing the name "Decision Analysis" (Raiffa, 1968).

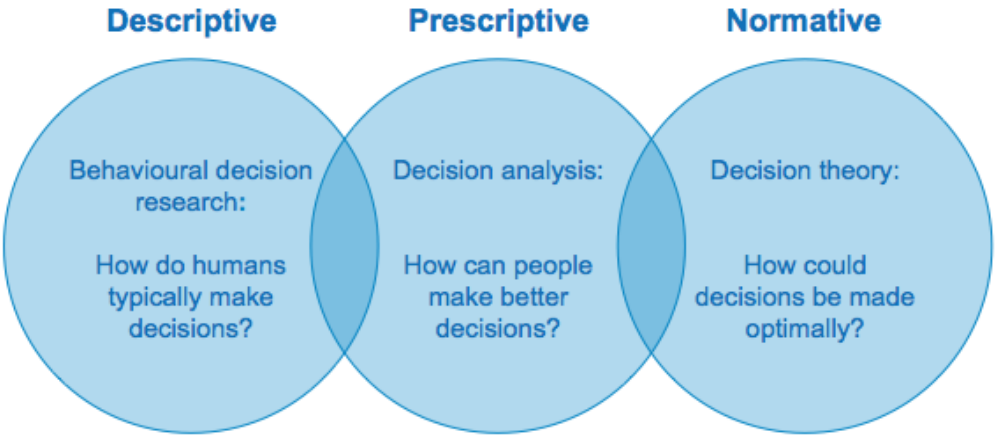


Figure 1-1 Three research foci of decision science (cf. D. E. Bell et al., 1988)

Prevalent research on DMC of individuals combines two of the above described decision science disciplines: behavioural decision research and decision theory. Behavioural decision research detects significant deviations of actual human decision-making behaviour from normative standards (e.g. Bruine de Bruin, Parker, & Fischhoff, 2007). This stream of research provides evidence that people differ within their decision-making behaviour (e.g. Slovic, 1962; Weber, Blais, & Betz, 2002), and perform differently on various decision-making tasks (e.g. Einhorn, 1970; Finucane & Lees, 2005; Finucane & Slovic, 2002; Hunt, Krzystofiak, Meindl, & Yousry, 1989). More recent research has begun to examine DMC as a psychological construct[5] (e.g. Stanovich & West, 2000) and analyse correlations between diverse decision-making tasks, or between single dimensions of DMC. As one of the first investigations, the work of Stanovich and West (2000) describes decision-making performance as a higher-level measure[6] with different assessable correlated dimensions. This direction of research paves the way for measuring DMC or aspects of it and consequently for an evidence-based definition.

---

[5] A construct (Cronbach & Meehl, 1955) is a term from scientific theory, especially in psychology. It is an explanatory variable, which is not directly observable. That is why psychologists often speak of a latent construct, for instance intelligence or motivation (Bortz & Döring, 2005).

[6] As section 2.4 describes, a good decision-making process consists of several dimensions, which will be aggregated into the higher-level measure of DAC.

Promising approaches of defining and measuring DMC include the studies of Parker and Fischhoff (2005) and Bruine de Bruin et al. (2007), which present DMC as a construct consisting of seven dimensions, derived mostly from behavioural decision research. The authors have created two performance tests; one for youths and one for adults. Both allow for measuring an individual's performance on seven dimensions of DMC (cf. Table 2-1, p. 37).

Most of the tasks included in the two tests examine the ability of decision-makers[7] to deal with different decision biases. The authors assume that individuals with poor DMC, on average, perform worse on the test and experience more negative decision outcomes[8] in the long run than those who show sound DMC. An exploratory factor[9] analysis of the performance test *Adult Decision Making Competence* index score (henceforth: A-DMC; Bruine de Bruin et al., 2007; cf. section 2.2) provides a one-factor model. It shows acceptable to good statistical properties in terms of significant test-retest reliability and internal consistencies. The work of Bruine de Bruin, Parker and Fischhoff "show[s] promise for the development of a normed psychometric test of" DMC (2007, p. 948), which serves to objectively assess humans' DMC.

Including Bruine de Bruin et al.'s approach, the few existing studies of DMC as a psychological construct primarily focus on the influence of biases on human decision-making behaviour. Data on the validity and reliability of DMC as a psychological construct is rare (e.g. Bruine de Bruin et al., 2007; Bruine de Bruin, Parker, & Fischhoff, 2012; Finucane & Gullion, 2010; Parker & Fischhoff, 2005).

Thus, the present work focuses on the decision-analytical perspective of individual decision-making, instead of the up to now traditional view on DMC as a construct based on the behavioural biases perspective. The present Ph.D. research aims to provide a decision-analytical definition of DMC and, on this theoretical basis, contrive a psychometric test that allows an individual's performance to be measured on the various dimensions of the so-called Decision-Analytical Competence (henceforth: DAC). DAC could be seen as one aspect of DMC or the combination of various analytical abilities of

[7] A decision-maker is "an individual who has the power to commit the resources" (J. E. Matheson & Howard, 1968).

[8] "[N]egative decision outcomes [are] sampled across a wide variety of domains and varying in severity (e.g., threw out food or groceries you had bought, got divorced, had a mortgage or loan foreclosed)" (Bruine de Bruin et al., 2007, p. 943).

[9] In statistics, the exploratory factor analysis examines correlating variables in terms of a hypothetical higher-level measure – the factor (Bortz & Döring, 2005).

the general construct of DMC. The following section explains the underlying research questions and specifies the research objectives and aims.

1.2 RESEARCH QUESTIONS & OBJECTIVES

This Ph.D. research uses a framework that is based on theoretical decision-analytical dimensions of a sound decision-making process, and aims to provide a valid and reliable psychometric performance test for individual DAC. It encompasses the following research questions and objectives (cf. Table 1-1):

Table 1-1 Research questions and objectives

Research question	Research objective
Q_1: What do most advanced approaches of defining and measuring DMC neglect?	O_1: Identify the strengths and weaknesses of prevalent definitions and measurements of DMC and analyse gaps for improvement.
Q_2: How can DAC be defined as a psychological construct, consisting of a concrete number of measurable decision-analytical dimensions?	O_2: Conceptualise DAC as a psychological construct, consisting of a concrete number of measurable dimensions according to the theory of decision analysis.
Q_3: What are appropriate items to operationalize the different theoretical dimensions of DAC?	O_3: Develop items to operationalize the various theoretical dimensions of DAC and construct a psychometric performance test to measure DAC.
Q_4: How well does the psychometric DAC performance test perform in terms of statistical goodness criteria according to its empirical evidence for sound decision-making measured by a set of appropriate criteria for validation?	O_4: Examine the proposed psychometric DAC performance test in terms of its reliability and validity according to its empirical evidence for sound decision-making as measured by a set of appropriate criteria for validation.

Note. Q = Research question; O = Research objective.

In the present research, decision analysis builds the research context, DAC the research objective, and performance testing the research method (cf. Figure 1-2).

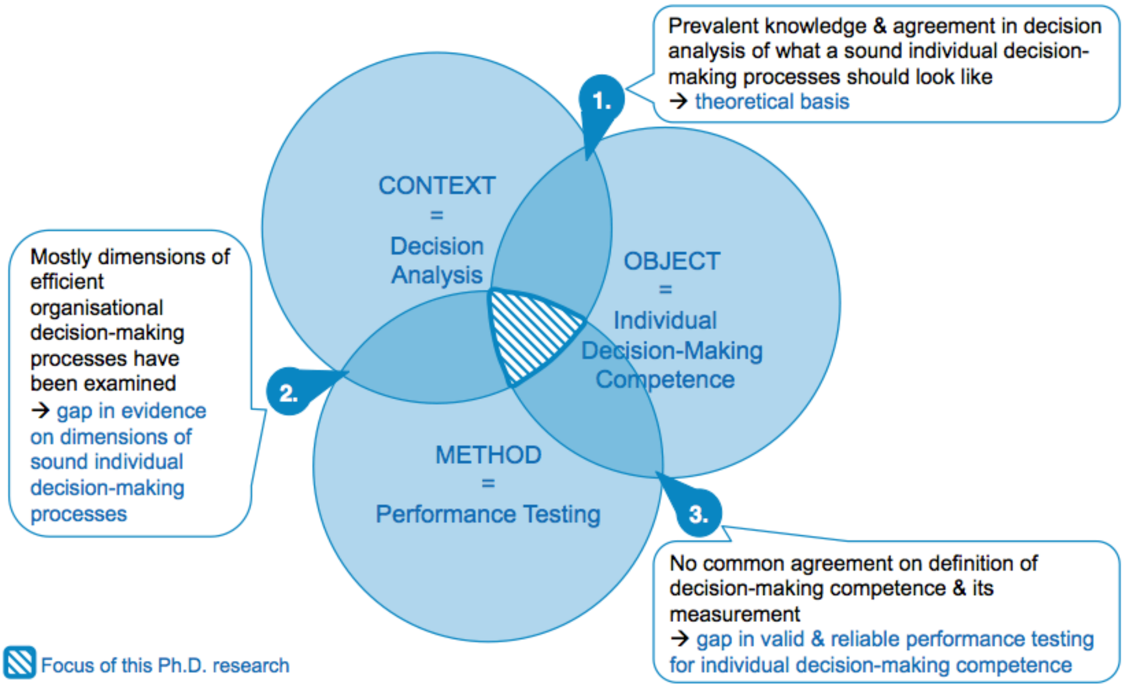


Figure 1-2 Three research-leading reasons for this Ph.D. research

Besides the lack of public attention to the importance of DMC, there are three research-leading reasons, illustrated by the overlaps of context, objective, and method in Figure 1-2, and corresponding aims for undertaking the proposed study:

First, the decision analysis literature provides a series of interrelated steps for making good decisions (Mellers & Locke, 2007). Researchers and practitioners from different traditions present similar dimensions of an analytical decision-making process (cf. section 2.4). Following this, the present research aims to build on the prevalent knowledge and partial agreement on potential dimensions of a good analytical decision-making process as a theoretical basis for the development of a measuring and assessing instrument for DAC[10].

Second, so far decision-analytical research has mostly concentrated on examining organisational decision-making processes in order to understand how desired decision outcomes, such as increasing revenue or market share, can be achieved in the long run (e.g. D. Matheson & Matheson, 1998). In comparison, little research effort has been undertaken to expose empirical evidence of the success factors for individual decision-

[10] Within this thesis the abbreviation "DMC" refers to decision-making competence (research) in general. The abbreviation "DAC" refers to the research objective and therefore the main focus of this thesis. While the construct of DMC, used in the thesis, describes the general construct, which explains how well people are able to solve decision problems successfully, ranging from small every-day choices to rare and complex decisions, DAC refers to an aggregated composition of analytical abilities necessary to deal with complex decision problems.

making processes – even though it is possible that better DAC of individuals would also support the improvement of organisational decision-making processes over time due to more highly-skilled employees (Hoffman, Woehr, Maldagen-Youngjohn, & Lyons, 2011). Thus, this Ph.D. research aims to contribute to filling this gap by focusing on the DMC of the individual.

Third, DMC has been examined in many studies, mostly in behavioural psychology and behavioural decision research. The literature presents several definitions of DMC and therefore several different testing procedures. The most common approach is the self-rating procedure (e.g. Mann, Burnett, Radford, & Ford, 1997; Siebert & Kunz, 2016). Given their limited validity in comparison to performance tests (e.g. Nisbett & Wilson, 1977), the current research aims to provide a psychometric performance test on DAC, which meets statistical properties such as validity and reliability.

In general, there are three reasons, illustrated in Figure 1-3, why providing a measurement instrument for DAC would be of interest. First, it would allow an individual's DAC to be measured on various dimensions. Hence, a person interested in their test score could receive detailed feedback about which dimensions of DAC are already well evolved and which dimensions need more training. Second, it would allow for differentiating between individuals based on their quality of DAC. So, if a company wants to recruit a new employee with good DMC, it could embed the test in an assessment centre and objectively choose the candidate with pertinent testing results[11]. Third, it would allow pre-post-tests to be conducted. Thus, class teachers or lecturers would be able to test students before and after a course in analytical decision-making or related subjects, compare the results of the two testing times and evaluate their impact.

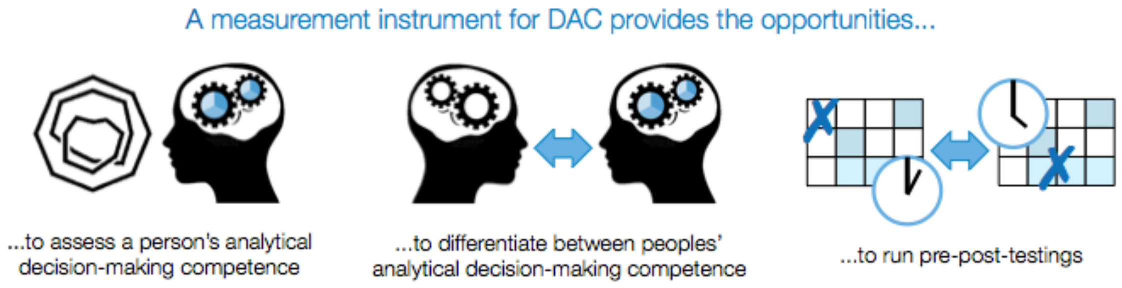


Figure 1-3 Why measuring DAC is of interest

Summarising, this research aims to investigate the psychological construct of DAC by describing a higher-level measure, which consists of different measureable

[11] The example refers to a fictive situation, in which DMC is the only relevant criterion for the HR department.

dimensions. The main goal is to provide a valid and reliable psychometric performance test for assessing DAC on the basis of decision-analytical insights. Empirical evidence for the correlation with desired decision outcomes will be measured by an appropriate set of different variables reflecting, for instance, satisfaction with already-made decisions or a vigilant decision-making style.

Gaining more insights on how to measure and assess DAC could help to define and establish DAC as a psychological construct. It could attract the public's attention, as has happened before to the psychological construct of intelligence during its first psychometric tests (e.g. Wechsler Adult Intelligence Scale, cf. Wechsler, 1939). In this respect, DAC deserves serious investigation. The next section presents the outline of this Ph.D. thesis.

1.3 OUTLINE

Following this introduction, chapter 2 presents the literature review on DMC. It delimits the research domain including relevant terms, and gives an overview of the status quo of defining and measuring DMC. Limitations of the existing research are pointed out, ideas for further research are described in greater detail, a decision-analytical consideration of the construct of DMC is introduced, and the theoretical compliance on DAC-dimensions is depicted.

Chapter 3 gives a brief overview of the research methodology. It lists the research hypotheses, explains the procedure of psychometric test development and illustrates the arguments for the specific methodological choices that have been made. Furthermore, the chapter also presents the target group and outlines the strategy for data collection.

Chapter 4 focuses on the development of the DAC-test. It presents the thoughts taken into account to construct and set up the test and explicates how each DAC-dimension has been operationalized. Additionally, the chapter presents the socio-demographics and the results of item and distributional analysis of the pre-testing.

Chapter 5 introduces and argues the selected criteria for validation. Therefore, the psychological constructs are defined and the theoretical hypotheses about their relationship with DAC are presented. Furthermore, this chapter constitutes the results of the main testing according to reliability and validity of the DAC-test. The parameters for objectivity are also argued.

Chapter 6 focuses on the interpretation and discussion of the results. It answers the research hypotheses and compares the present findings with one of the most advanced approaches to measure DMC. It addresses the scientific contribution of the work and its potential for practical utilisation. Additionally, the limitations of the research results are addressed and potential avenues for further research illustrated.

Chapter 7 concludes the Ph.D. research on DAC and its test development and highlights why this research approach should be of interest to the scientific field of decision science and also society.

2 DOMAIN DEFINITION AND LITERATURE REVIEW

The first step in the process of answering the research questions and addressing the research objectives presented in section 1.2 is to theoretically capture the construct of DMC from different perspectives by giving an overview of its various definitions and diverse measurement approaches. This review of existing literature in the area of decision science is presented in order to allow a systematic deduction of the definition of DAC to take place.

The following sections of this chapter outline a literature review for this Ph.D. research. While section 2.1 presents definitions and clarifications, section 2.2 reviews the current status of defining and measuring DMC in the decision science literature. Referring to this status quo, section 2.3 exposes the limitation of well-advanced measuring approaches and introduces potential for improvement. Section 2.4 argues how a decision-analytical avenue of research can help to fill the gap presented in the previous section. It presents an overview of the theoretical dimensions of DAC, and summarises the theoretical agreement on dimensions of DAC presented by the decision analysis literature. Closing the literature review, section 2.5 summarises the theoretical consensus on the concept of DAC over different decision scientific research groups.

2.1 THE CONCEPT OF "DECIDING"

At the beginning of this literature review, it is necessary to define what is meant by the term *decision* and the corresponding terms used in this research context, and to concretely specify which type of decision the current research is focusing on.

The term *decision* used in this Ph.D. research refers to the definition of Mintzberg, Raisinghani, and Theoret (1976). They "define ... a decision as a specific commitment to action (usually a commitment of resources) and a decision process as a set of actions and dynamic factors that begins with the identification of a stimulus for action and ends with the specific commitment to action." (p. 246) The definition of Mintzberg and colleagues emphasises that a decision not only defines a state, but also invokes a process with a set of cognitive performances, starting with the recognition of a decision problem and closing with a commitment to action.

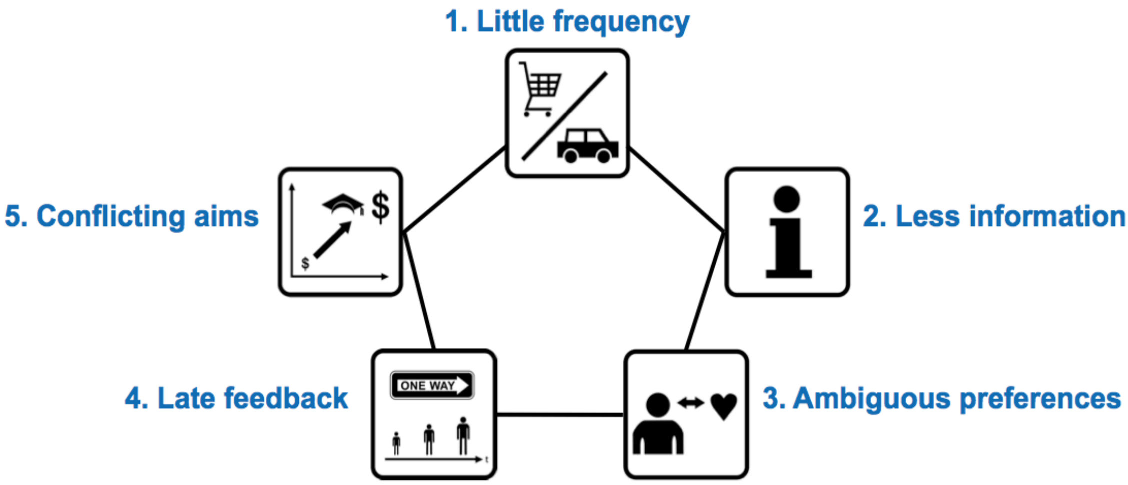


Figure 2-1 Attributes of complex decisions

The present research focus lies on complex and demanding decisions, which represent an intense cognitive challenge (e.g. Simon, 1978). Thaler and Sunstein (2008) describe five attributes (cf. Figure 2-1) that make *complex decisions* appear particularly challenging for decision-makers and explain why these kinds of decisions are so difficult to deal with.

First of all, complex decisions do not occur often. They are rare. This circumstance harbours two challenges: In comparison to minor frequent decisions, as for instance what to get from the grocery, for which it is possible to quickly apply rules of thumb (e.g. Gigerenzer & Gaissmaier, 2011; Gigerenzer, Todd, & the ABC Research Group, 1999), decision-makers have no personal or just limited experiences with a particular complex decision type. In the latter case, decision-makers cannot compare the present situation with previous decisions and therefore have to develop new strategies, which takes time and might involve trial and error. Other people – potential role models – also experience a particular complex decision, such as deciding which vehicle to buy, only a few times in their lives. Thus, there are not as many chances to observe and learn from others. Observational learning (Bandura & Walters, 1963), the most common way of human learning, cannot take effect as it does for minor decisions, which take place every day.

Second, due to their intricacy complex decisions are difficult to handle. To appropriately deal with them, it is practical “to simplify a complex decision environment to a manageable size for analysis.” (W. Edwards et al., 2007, p. 5) Decision-makers need information on the decision’s details such as circumstances, alternatives, corresponding probabilities, potential consequences, etc. (e.g. Howard, 2007; von Winterfeld & Edwards, 2007). Particularly for complex and rare decisions, people often do not have access to all the required information that is needed to make a good decision. As a result, missing information and uncertainty lead to a fragmentary picture of the real world problem and bedevil the problem analysis.

Third, in many situations humans are not able to envision their preferences adequately (e.g. Hsee & Hastie, 2006). "Decision-makers are considerably deficient in utilizing personal knowledge and values to form objectives for the decisions they face" (Bond et al., 2008, p. 56). As each of the objectives can be seen as a desire that the decision-maker wants to be fulfilled, the articulation of one's individual definite and exhaustive set of objectives is indispensable to achieve one's desirable decision outcome.

Fourth, feedback on a particular decision process and its outcome, from which decision-makers could learn, usually does not appear immediately after the complex decision has been made. Especially, the achievement or realisation of *fundamental objectives* in comparison to *means objectives* (Keeney, 1992)[12] can take several years for this kind of decision. This circumstance makes it difficult to improve one's decision-making process by learning from previous decision-making experiences. Failures in the decision-making process might not be identified and corrected, which reduces the learning effect to improve the quality of decision-makers' way of dealing with complex decision problems.

Fifth, the costs of some decisions, such as time, rational thought, or money do immediately incur, while the benefits of those decisions take much longer to appear. In this sense, decision-makers are confronted with conflicting facts such as immediate costs on the one hand and mid- and long-term benefits on the other. Thus, in comparison to minor decisions it is more challenging for decision-makers to cognitively balance the costs or effort of a decision in relation to its benefit. In other words, decision-makers experience problems by making trade-offs when no alternative is dominant over all other alternatives on all objectives. "The most difficult choices occur when there are negative correlations among the values of the attributes across the alternatives, forcing us to make difficult trade-offs" (Hastie & Dawes, 2010, p. 218). So for instance in the case of choosing the fitting education for one's future, objectives like attractiveness to potential employers, costs and duration, depth of education may conflict. Consequently, this occurrence makes it harder to follow through with tough decisions and transfer the cognitive decision into corresponding action.

[12] Decision objectives can be divided into two categories: fundamental objectives and means objectives. Whereas fundamental objectives describe "desires, which are an end in themselves for the situation at hand, and accomplishment of a fundamental objective provides direct utility to the individual. Conversely, means objectives provide utility by facilitating the achievement of other objectives. In general, the set of fundamental objectives provides the basis to evaluate various alternatives, whereas means objectives help to stimulate the generation of alternatives." (Bond et al., 2008, p. 58/59)

In summary, complex decisions are especially challenging as the decision-makers have limited experience with similar situations that they can draw upon and less information to base their arguments upon. In this context, it is difficult for decision-makers to envision one's different, partly conflicting objectives and make trade-offs. Also, feedback on how well a decision was made comes at a later time, which hinders the following-through with the decisions.

To support people dealing with those complex and demanding decisions, behavioural psychologists and decision scientists have studied human decision-making behaviour in different decision situations for decades. Consequently, they have developed strategies to cope with decision complexity.

As mentioned earlier, one approach created in the late 1960s (Howard, 1966; Raiffa, 1968) to improve decision-making processes is called *decision analysis*. "It is ... a logical balancing of the factors that influence a decision. ... discipline has two interesting aspects. First, it is a language and philosophy for decision-making. It is a way to talk about the decision-making process ... [and second,] it is a logical and quantitative procedure". (Howard, 1973, p. 64) Being aware of human cognitive limitations, which are addressed by behavioural decision research, decision analysis makes use of normative models (D. E. Bell et al., 1988). This research field aims to enhance the decision-making processes of decision-makers by "[c]losing critical gaps between the normative ideal and the descriptive reality." (Fischhoff, 2008, p. 14) Decision-analytical scientists and practitioners see the additional benefit of decision analysis in presenting and communicating uncertainty in a decision as one of the most important advantages (e.g. Howard, 1980; Matheson & Howard, 1968), as well as modelling multiple objectives (Keeney & Raiffa, 1993). "The strength of decision analysis ... lies in increasing the logical quality of decision-making." (J. E. Matheson & Howard, 1968, p. 4) It means dividing the elements of a decision into smaller dimensions, analysing the situation and assembling those smaller problems back to a manageable decision problem. Decision analysis presents a decision problem in an appropriate way, which is easier to comprehend for decision-makers (W. Edwards, Miles Jr., et al., 2007). Bazerman and Moore (2009, p. 181) state: "Since we do not make optimal decisions intuitively or automatically, when decision quality really matters, it makes sense to rely on procedures that can help direct us toward more optimal decisions". Consequently, the discipline of decision analysis aims to support people in making better decisions and thereby, helps to improve the quality of the decision-making process (e.g. Schilling, Oeser, & Schaub, 2007).

In this context, it is important to point out the distinction between a *good decision* or decision process and a *good decision outcome*.

"A good decision is a logical decision – one that is based on the uncertainties confronted by and on the values and preferences of decision-makers. But a good decision must not inevitably in any case lead to a good decision outcome. A good decision outcome is one that is profitable or otherwise highly valued" (Howard, 2008, p. 98). For example, parents teach their children that when standing in front of a pedestrian light one should walk while green and wait while red. This rule is supposed to be a good decision, as the chance of getting safely to the other side is much higher when walking when green. However, quite a lot of people decide to cross the street when red and reach the opposite pavement safely. From a decision-analytical point of view one would call this choice a poor decision even though the outcome has turned out well. In contrast, it is also possible to cross the road while green and be hit by a car. In this case, decision science would talk about a sound decision with a bad outcome.

Consequently, "[g]ood outcomes are what we desire, whereas good decisions are what we can do to maximize the likelihood of having good outcomes" (McNamee & Celona, 2005, p. 2). Ward Edwards (1954, p. 7) explains: "[a]ll real decisions are made under uncertainty. A decision is therefore a bet, and evaluating it as good or not must depend on the stakes and the odds, not on the outcome." Thus, the quality of a decision cannot be measured by its outcome (e.g. Hastie & Dawes, 2010; Keelin, Schoemaker, & Spetzler, 2009), since we cannot control all factors that influence the decision's outcome. "[I]t is the potential outcomes, their probabilities, and their values to the decision-maker at the time the decision is made that lead us to judge a particular choice to be wise or foolish." (Hastie & Dawes, 2010, p. 16) This conceptualisation emphasises the importance of a sound decision-making process and why it is crucial to examine an individual's DAC.

After elucidating the decision concept by defining research relevant terms such as *complex decisions, decision analysis, good decision, and good decision outcome*, the following section serves to theoretically define the domain of this Ph.D. research in detail by a literature review of DMC. It presents a distinct definition of DMC as a psychological construct and outlines how it is measured up to this point.

2.2 CURRENT STATUS OF DEFINING & MEASURING DECISION-MAKING COMPETENCE

In general, the term *measurement* describes "the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to characterize the attributes by clearly defined rules." (Fenton & Whitty, 1995, p. 6) In social research - the methodological background of this research - the main “goals ... are to understand, explain, and make inferences about social phenomena” (Agresti & Finlay, 1997, p. 12). Therefore, data has to be gathered. In psychology the researcher has to translate their imaginary idea of a psychological construct into useful data through measurement. A *construct* – the object to be examined - is not empirically recognisable and not an observable issue of a scientific theory. Hence, the researcher needs indicators to make such a construct measurable. That is why constructs are also called *latent constructs*[13.] The process of making latent constructs measurable is called *operationalization* (Cronbach & Meehl, 1955; cf. section 4.1).

When referring to the term *competence* in this research the “capacity to make a rational or intelligent judgment” is meant (Tancredi, 1982, p. 53). Finucane and Lees (2005, p. 6) explain: “the term ‘competence’ has most commonly referred to diagnostic criteria for determining an individual’s ability to make important life decisions”. Thus, within this Ph.D. research the term *competence* per se is understood as the possession of a combination of different abilities or capabilities of an individual, which can be learned, trained, and improved.

Before introducing a decision-analytical avenue to this field of research in section 2.4, the construct of DMC is defined and its measures reviewed. As Appelt et al. (2011) and their widespread and continuously evolving online database[14] “highlight[s] … the most common categories of individual difference measures used in judgment and decision-making research” (p. 253), this research does not intend to give an all-embracing review of the complete set of different approaches to define and measure DMC. Moreover, this research aims to examine the most advanced approach of defining and measuring DMC from a methodological background, to analyse its strengths and weaknesses, and to make suggestions for improvement by not only presenting a theoretical decision-analytical avenue to DMC, but also an evidence-based testing instrument to measure it. Following, it is indented to theoretically define and practically measure DAC.

To suggest a definition for the construct (DMC) and present a testing instrument, several studies of DMC and its dimensions have been undertaken. While some studies

---

[13] One well-known example for a psychological latent construct is the intelligence and its indicator IQ, which is measured by psychometric tests such as the Wechsler Adult Intelligence Scale-Revised (Wechsler, 2008).

[14] Decision Making Individual Differences Inventory (www.sjdm.org/dmidi/)

provide a literature overview of prevalent approaches to define and measure the construct according to different content foci and measurement tools (e.g. Appelt et al., 2011; Finucane & Lees, 2005; Mann, Harmoni, & Power, 1989), other studies present one concrete definition in combination with a testing instrument (e.g. Bruine de Bruin et al., 2007; Mann, Burnett, Radford, & Ford, 1997; Parker & Fischhoff, 2005). However, so far "[n]o single set of criteria has been agreed on for determining DMC." (Finucane & Gullion, 2010, p. 20) The multiplicity of research avenues related to the construct results in DMC appearing as a multidimensional concept (Finucane & Gullion, 2010), requiring several, mostly cognitive, competences.

In this respect, several studies make suggestions for a set of required skills to define DMC: In 1977 Janis and Mann presented the *vigilant decision-maker* "as a highly competent person who thoroughly canvasses a wide range of alternative courses of action, surveys a full range of objectives and values implicated by the choice, carefully weighs the positive and negative consequences that could flow from each alternative, intensively searches for new information, incorporates new information even when it is unpleasant, and plans for the implementation of the decision." (Mann et al., 1989, p. 266/267) In comparison, the analysis of important aspects of good decision-making of Mann et al. (1989) resulted in a list of nine relevant dimensions: choice, comprehension, creativity, compromise, consequentiality, correctness, credibility, consistency, and commitment. With this approach, the authors extend Janis' and Mann's concept of vigilance in decision-making and Ross' (1981) five skills of information processing[15]. They describe a set of cognitive processes including the search for and the process of information, problem solving, judgement, learning and memory. Furthermore, a respectable number of approaches from different decision science groups (cf. section 2.4) introduce concrete decision-analytical skills for making sound decisions (e.g. Hammond et al., 1999; Howard, 2007; Keelin et al., 2009).

All of those approaches describe cognitive skills or steps of action for good decision-making. To date, none of those approaches have been translated into a psychometric test instrument that assesses a person's current performance capabilities according to the decision analytic dimensions.

Historically, two methods have been used to measure different decision relevant

[15] Five skills of information processing: identifying alternative courses of action, identifying appropriate objectives to assess options, assessing options by objectives, summarising information on options, and checking the results of one's analysis

constructs, such as decision-making processes, behaviour, attitude, or competence (Finucane & Gullion, 2010). The first method, using verbal as well as written assessment, is the questionnaire. The second method, observing a decision-making process by measuring performance, is the performance test. In a narrow sense, a *test*[16] describes a procedure that measures a specific performance (e.g. Zimbardo & Gerrig, 1996). In comparison, a questionnaire is a self-report of a subject, in which the subject explains their opinion, attitude, and interests or gives information about their personality traits or background information (e.g. Zimbardo & Gerrig, 1996). In contrast to self-reports, performance tests provide several advantages (cf. section 3.2). Nevertheless, the most common way to measure DMC so far is the procedure of self-rating. This approach requires decision-makers to self-report their own decision-making behaviour. Well-known examples of questionnaires used for assessing individuals' DMC are: the *Flinders Adolescents Decision Making Questionnaire* assessing decision self-esteem, decision vigilance, tendency for hasty and impulsive choice, tendency of decision avoidance, tendency to apathy and non-involvement (Mann, Harmoni, Power, Beswick, & Ormond, 1988), the *Virgil Questionnaire* assessing "goals clarification, generation of options, search for facts, consideration of effects and review of action" (Mann, Harmoni, Power, et al., 1988, p. 163), and the *Melbourne Decision Making Questionnaire* assessing tendency of unconflicted adherence, tendency of unconflicted change, tendency of defence avoidance, tendency for hypervigilance, and the tendency for decision vigilance (Mann et al., 1997).

A more recent way to assess DMC[17] is the ability test, which assesses a person's maximum performance of decision-making in a concrete situation. Often "developed and tested on impaired or ill/hospitalized persons" (Finucane & Gullion, 2010, p. 19), these ability tests measure the competence to decide about medical treatments or medication. Only a few studies focus on everyday DMC. In this respect, one very advanced approach to measuring DMC is the *Adult Decision-Making Competence* index score of Bruine de Bruin, Parker and Fischhoff (2007). For developing the A-DMC, the authors selected seven decision-making tasks with a pedigree in the experimental literature to "reflect the

[16] Even in psychological literature, often the term "test" is used for any measurement instrument, independent of talking about a questionnaire (=self-rating procedure) or a performance test. Within this Ph.D. research, the term "test" is only used to describe a psychometric performance test.

[17] Existing research usually uses the term decision-making competence for designed measures mostly referring to behavioural aspects of decision-making. This ascription might be misleading as one could assume that the construct of decision-making competence is all-embracing, e.g. behavioural and decision-analytical facets of the constructs.

traditional normative approach to decision-making competence" (p.949). Participants' performances across the seven tasks are aggregated into the A-DMC index score. The seven decision-making tasks are: *resistance to framing, recognising social norms, under/overconfidence, applying decision rules, consistency in risk perception, resistance to sunk costs, and path independence.* Table 2-1 gives an overview of what each A-DMC index score task aims to measure.

The A-DMC is based on the work of Parker & Fischhoff (2005), who introduced the *Youth Decision-Making Competence* index score (henceforth: Y-DMC). The Y-DMC applied the same decision-making tasks as the A-DMC, but aligned and presented them to adolescents in a long-term study. Bruine de Bruin et al. (2007) adapted their decision-making test in order to fit adults' background and intended to increase its statistical quality criteria. However, the Y-DMC is now seen as a prototype (A. M. Parker personal communication, September 1, 2012), since the psychometric properties of the A-DMC are better than the properties of the Y-DMC. In comparison to the results of the Y-DMC with 25.1% of explained variance, Bruine de Bruin and colleagues were able to increase the explained variance. An exploratory factor analysis of the performance test A-DMC (Bruine de Bruin et al., 2007) provided a one-factor model, which explains 30.1%[18] of the total variance (p. 944). Additionally, the A-DMC showed acceptable to good statistical properties for significant test-retest reliability[19] at the $p < .001$-level. Furthermore, four out of its seven dimensions met the threshold of Cronbach's alpha (Cronbach, 1951) with $\alpha \geq .70$ (e.g. Bland & Altman, 1997) – a theoretical equivalent to the mean of all split-half correlations measuring internal consistencies[20].

Bruine de Bruin et al. (2007) were able to prove that the tasks show "good internal consistency, test-retest reliability, and validity – as seen in significant, predicted correlations with real-world decision outcomes" (p. 949) by asking for self-reported experiences of real-world outcomes measured by the developed and parallel tested *Decision Outcome Inventory* (henceforth: DOI). The DOI measures success of individual decision-making by self-reporting in terms of avoiding specific negative decision

[18] 30.1% of variance explanation means that almost one third of the shown and measured behaviour (good decision outcomes measured by the Decision Outcome Inventory (Bruine de Bruin et al., 2007)) could be explained/predicted by the measured dimensions/items.

[19] To estimate the retest reliability the same sample group of examinees gets tested two times and the correlation of both test scores gets calculated. The retest reliability of a test is high, if two testings with the same test at different time points highly correlate (Bortz & Döring, 2005).

[20] If a test contains homogeneous items, each item can be seen as part of the test to measure the trait. To calculate the internal consistency Cronbach's alpha gets calculated. The higher the correlation of the items on average, the higher the internal consistency (Bortz & Döring, 2005).

outcomes; for instance, "[i]n the last 10 years, have you ever... missed a flight [item 10b]" or "[b]een in a jail cell overnight for any reason [item 29]" (Bruine de Bruin et al., 2007, p. 956). Reduced from a much longer list of negative decision outcomes, 41 negative decision outcomes with good internal consistency have been selected for their scale. By creating the DOI it was assumed that individuals with poor DMC perform worse on the A-DMC and experience more negative life events, assessed by the DOI, in comparison to those who show good DMC.

Table 2-1 The seven A-DMC index score tasks

A-DMC Task	Issue of measurement
Resistance to framing	Assessing the tendency to be affected by irrelevant variations in problem description
Recognising social norms	Assessing participants' identification of values by their attention to a common value
Under/overconfidence	Assessing peoples' ability to recognise the extent of their knowledge
Applying decision rules	Assessing peoples' ability to follow given decision instructions
Consistency in risk perception	Assessing individuals' ability to follow probability rules
Resistance to sunk costs	Assessing the ability to ignore prior investments when making decisions
Path independence	Assessing peoples' consistency of participants' choices in games of chance

Note. A-DMC = Adult Decision-Making Competence index score (Bruine de Bruin et al., 2007).

In summary, over the last decades the literature has presented a variety of approaches defining and assessing DMC and decision-making behaviour. So far, no overall agreement exists, neither on the various dimensions of DMC nor on its measurement instrument. More recent research focuses on procedures of performance measurement. The most advanced approach, in terms of good statistical properties, seems to be the A-DMC index score (Bruine de Bruin et al., 2007) – a test that focuses on insights from behavioural and normative decision science. Concentrating on the A-DMC index score and its forerunner the Y-DMC (Parker & Fischhoff, 2005), section 2.3 describes the limitations of common instruments and potential for improvement.

2.3 LIMITATIONS & POTENTIAL FOR IMPROVEMENT OF EXISTING MEASUREMENTS

As outlined above (cf. section 2.2), the two most advanced approaches in the research field of decision theory are the Y-DMC and the A-DMC index scores. In developing these psychometric performance tests, Parker & Fischhoff (2005) and Bruine de Bruin et al. (2007) followed the suggestions of various scientists (e.g. Milkman et al., 2009; Stanovich & West, 2000) and examined DMC as a higher-level measure by dividing it into several correlated dimensions. Especially the more evolved A-DMC test presents multiple advantages compared with former approaches of measuring DMC by self-reporting procedures. It is performance-based and represents DMC as an underlying construct or competence with sub-scales or abilities. Additionally, the A-DMC provides acceptable to good psychometric properties in terms of its validity and reliability. Its tasks are based on existing studies, which proves their internal consistency. The A-DMC has been applied in several studies (e.g. Bruine de Bruin, Parker, & Fischhoff, 2012; Jacobson et al., 2012; Parker, Bruine de Bruin, & Fischhoff, 2015; Szanto et al., 2015) and holds the potential to become a standard test instrument for DMC.

Nevertheless, the A-DMC shows limitations, which the present research intends to address partially.

First, paying attention to the A-DMC and its sub-scales, it becomes apparent that for the Y-DMC and A-DMC the authors selected seven sub-scales to “fit the theoretical categorization of normative decision-making skills” (Bruine de Bruin et al., 2012, p. 940)[21]. Thus, both tests focus on insights from behavioural decision research and decision theory. Out of the seven sub-scales of the A-DMC (cf. Table 2-1) none of them are dedicated to the application of decision-analytical rules. Hence, the A-DMC measures the quality of a person’s DMC by primarily testing for decision-makers’ resistance to decision biases. It is proven that biases have a great influence on human decision-making and judgement (e.g. M. Weber & Borcherding, 1993) and it is assumed that the better the decision-maker’s competence to deal with decision biases, the better their decision-making behaviour (Larrick, 2004). Nevertheless, for making sound decisions and, hence, potentially achieving desired decision outcomes in the long term (e.g. W. Edwards, 1984; Howard, 1973; Keeney, 2008; Keren & Bruine de Bruin, 2003; Larrick et al., 1993) or increasing the likelihood of experiencing sound decision outcomes (W. Edwards, 1984; Keren & Bruine de Bruin, 2003), the decision process has to be more than just free of decision biases (e.g. Baron, 2008). For sound decision-making in complex decision situations, decision-makers have to complete several decision-analytical steps in a

[21] For the Y-DMC Parker and Fischhoff (2005) have selected tasks capturing each of the four fundamental decision-making skills presented by Edwards (1954) or Raiffa (1968): assessing beliefs, assessing values, combining beliefs and values in order to identify choices, and having a meta-cognitive understanding of one’s abilities.

decision process (cf. section 2.4). The A-DMC defines DMC mostly by the absence of the biases' influence. It is developed from a behavioural avenue of decision research and thus, does not aim to cover decision-analytical insights or the execution of concrete decision-analytical steps.

Second, the sub-scale *applying decision rules* of the A-DMC is characterised by a certain degree of impreciseness, giving the impression, according to its name, that it refers to an analytical decision-making process but instead aims to capture people's ability to follow given instructions. Thus, it rather measures *test-taking skills* as for instance *following the instructions precisely* (cf. Pike, 1973; Rogers & Harley, 1999). Besides "the [actual] cognitive ability or basic skill the test is designed to measure ... [the test-taking skill describes] the ability to demonstrate cognitive ability or basic skill within the test situation" (Sabers, 1975, p. 7). Kettler, Braden, & Beddow (2011) define this kind of "test-wiseness [as] an access[22] skill for tests designed to measure academic achievement in a variety of content areas" (p. 148). In the context of performance testing, *access* is related to the opportunity of test participants to show their actual ability postulated by the test. The A-DMC sub-scale *applying decision rules* is operationalized by seven cases, for instance presenting a decision-maker who has to select a DVD player among five options according to decision-rule constraints (cf. Figure 2-2). Bruine de Bruin, et al. (2007) measure the quality of performance by the percentage of items for which the correct DVD players were chosen, based on the applicable decision rule. From a test developmental point of view, the main criticism for this A-DMC sub-scale is that the instruction (cf. Figure 2-2) already includes the solution to the tasks (Kline, 2015): By reading, understanding, and then precisely following the instructions, test participants can solve this kind of item. Hence, the A-DMC sub-scale *applying decision rules* could be seen as a proxy variable for test-taking skill or test wiseness.

[22] The better its access the greater the test's reliability and construct validity as "[b]oth reliability and construct validity are higher when the proportion of variance within a set of test scores that is construct relevant is maximized" (Kettler et al., 2011, p. 147).

→ LaToya only wants a DVD player that got a "very high" rating on reliability of brand.
Which **one** of the present DVD player would LaToya prefer?

DVD player	Picture quality	Sound quality	Programming options	Reliability of brand	Price
A	3	5	5	1	$369
B	1	2	1	1	$369
C	5	5	4	4	$369
D	5	3	4	2	$369
E	4	5	2	2	$369

Figure 2-2 Example item for "applying decision rules" (A-DMC; Bruine de Bruin, et al., 2007, p.954)

Third, even though the authors indicate that the Y-DMC as the first version of the A-DMC "capture[s] an overall picture of decision-making competence" (Parker & Fischhoff, 2005, p. 3) it does not become apparent to which decision type Y-DMC and A-DMC are tailored. When examining the chosen items of the tests, it is not clear whether the tests aim to focus more on fast and frugal decisions, or more on complex decision situations. This specification appears necessary, as it requires different strategies to effectively solve either type of decision. While successfully coping with minor or daily decisions calls for heuristic techniques (e.g. Todd & Gigerenzer, 2000), a sound analytical decision-making process is needed to deal with demanding and influential choices (e.g. Matheson & Howard, 1968).

Additionally, it is debatable whether content-disconnected tasks[23], (e.g. cf. Figure 2-3), which are picked to "fit the theoretical categorization of normative decision-making skills" (Bruine de Bruin et al., 2007, p. 940) from existing literature, provide testing data that is comparable and predictive to actual decision behaviour in real-life. A test is always just an attempt to reconstruct a real-world issue and create a measurement environment that is as close as possible to reality. However, in due consideration of test theoretical standards (e.g. Crocker & Algina, 2006) a test should aim to be the best approximation of the real-life issue it is referring to. The items chosen for the A-DMC do not represent problems decision makers usually face consecutively in their daily life. That is why the face validity for the measure does appear relatively low. The biggest challenge seems to be the

---

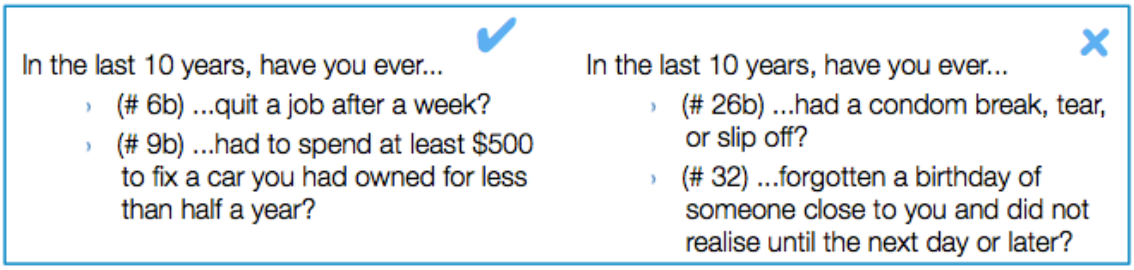
[23] With disconnected tasks items are meant that do not refer to one decision. To solve a real-life decision more than one skill would be needed at a time.

content validity of the A-DMC index, as it does not become clear how the authors have selected this specific set of decision tasks to capture the behavioural aspect of DMC.

Example item for *resistance to framing*
Imagine that recent evidence has shown that a pesticide is threatening the lives of 1,200 endangered animals. Two response options have been suggested: If Option A is used, 600 animals will be saved for sure. If Option B is used, there is a 75% chance that 800 animals will be saved and a 25% chance that no animals will be saved. Which option do you recommend to use?
Example item for *recognising social norms*
Out of 100 people your age, how many would say it is sometimes OK . . . . . . to steal under certain circumstances?
Example item for *resistance to sunk costs*
After a large meal at a restaurant, you order a big dessert with chocolate and ice cream. After a few bites, you find you are full and you would rather not eat any more of it. Would you be more likely to eat more or to stop eating it?

Figure 2-3 Example items for disconnected items of the A-DMC (Bruine de Bruin, et al., 2007, pp. 953-955

Fourth, even if the A-DMC and the DOI are significantly correlated with r_p = .29 (p < 0.001, Bruine de Bruin et al., 2007, p. 946), regarding some of the items (cf. Figure 2-4) of the DOI it is questionable whether the quality of DMC is linked with the tendency of decision outcomes. The criterion for validation – the DOI - aims to measure the success of individual decision-making in terms of avoiding specific negative decision outcomes (Bruine de Bruin et al., 2007). Reduced from an extensive list, 41 negative decision outcomes with good internal consistency have been selected by the authors. While some of the DOI items (cf. left column in Figure 2-4) plausibly ask for decision outcomes, which could have been influenced by a poor decision-making process, other items (cf. right column in Figure 2-4) are criticisable, as the connection between the appearance of the outcomes and the poor decision-making appears to be loose. Thus, it seems inexplicable why certain outcomes, such as condom break, tear, or slip off (cf. Figure 2-4), should be linked to poor decision-making as nowadays in the modern western world, people are obligated to have safe sex to protect themselves and others against sexually transmitted diseases. In other words, the evaluation of decision outcomes depends either on the perspective of the evaluator or on the situational context of the decision maker. This kind of subjectivity leads to a justifiable reservation towards the DOI as an appropriate and rigorous validation criterion. It is therefore desirable to find other criteria for validation to connect sound DMC to worthwhile decision outcomes.


In the last 10 years, have you ever...
- (# 6b) ...quit a job after a week?
- (# 9b) ...had to spend at least $500 to fix a car you had owned for less than half a year?

In the last 10 years, have you ever...
- (# 26b) ...had a condom break, tear, or slip off?
- (# 32) ...forgotten a birthday of someone close to you and did not realise until the next day or later?

Figure 2-4 Two example items of the Decision Outcome Inventory (Bruine de Bruin et al., 2007, p. 956)

In conclusion, the research on defining and operationalizing DMC has been broadened and improved within the past decades. More recent approaches have moved from self-rating procedures to psychometric performance tests, and so provide better psychometric properties. So far, these new approaches have focused mainly on insights from behavioural decision research and decision theory. However, these avenues do not consider decision-analytical knowledge.

For this reason, this Ph.D. research intends to capitalise on the strengths of two research directions and merge them. First, it proposes to build on the insights of decision analysis about analytical steps of sound individual decision-making. Second, it proposes to deploy the prevalent research on psychometric performance measures to guide the methodological considerations to reach better statistical properties, as valid and reliable performance-based measures of DMC are rare. In this respect, this research aims to present a new definition and a novel instrument to operationalize DMC, which presents the construct as a higher-level measure with several decision-analytical dimensions that correlate with each other.

2.4 DIMENSIONS OF DECISION-ANALYTICAL COMPETENCE

In this section, a decision-analytical avenue of research is considered to address the research gap presented in section 2.3. To give a visualised overview of the theoretical dimensions of DAC presented by decision-analytical literature a graphical representation, the *Process Cycle of Decision-Analytical Competence* (cf. Figure 2-5) is conceived. On this basis, a literature review of DAC is outlined.

The *Process Cycle of Decision-Analytical Competence* intends to illustrate cognitive motivational dimensions[24] of DAC[25]. The framework represents an individual's

---

[24] The framework does not claim to present a perfect and complete set of factors influencing a decision-making process. In fact it maps the cognitive and motivational components of decision-making. It omits the ability to deal with biases or volitional behavioural components necessary to transfer a decision into action.

decision-making process by a cycle, as decision-making can be seen as an on-going process, in which decision-makers try to find the best solution for their decision problem and think about the different facets of a decision. The consecutive nature of a decision-making process is that the cycle does not provide a starting or an ending point. Also, the literature does not clearly define where a decision-making process begins. The following remarks are ordered corresponding to a succession, which seemed reasonable for describing such a process.

The cycle consists of eight dimensions. All dimensions are drawn equally sized, as no dimension has a higher set value than another. In fact, Hammond et al. (1999) describe that a sound decision-making process can only be carried out if all dimensions have been addressed. To skip one of the steps could negatively affect the other dimensions and hence lower the chance of reaching the desired output and/or outcome (Mellers & Locke, 2007).

The eight dimensions of the *Process Cycle of Decision-Analytical Competence* are presented in three different colours to illustrate three phases of a decision-making process: *decision recognition*, *decision reasoning*, and *decision intention*. All three phases and their underlying facets represent cognitive abilities. From an *action theoretical* point of view they can be seen as motivational elements. In action theory, for instance Heckhausen (1989), there is differentiation between *motivation* and *volition*. Everything that contributes to forming an intention is called *motivational*. The *volitional* phase describes the momentum when the intention becomes implemented. In the context of the present research, the *Process Cycle of Decision-Analytical Competence* intends to capture the cognitive motivational phases of analytical decision-making[26] as the testing of volitional aspects in the context of complex decision-making requires action not reasoning, which would have exceeded the manageable frame conditions of the research project.

---

[25] Referring to decision-making competence in this research, a multi-dimensional concept is meant as various cognitive abilities for good decision-making are required (e.g. Baron, 2008). Comparable to similar studies in the research field of decision science, DMC is understood as an aggregated measure of a battery of different decision skills (e.g. Bruine de Bruin et al., 2007; Finucane & Gullion, 2010; Parker & Fischhoff, 2005). The DAC-dimensions are seen as some of those skills chosen to present the decision-analytical side of decision-making. Correspondingly, DAC can be understood as a "formative measure" (J. R. Edwards & Bagozzi, 2000) or an "index" (Streiner, 2003) specified by its separate dimensions.

[26] In section 2.4.3 it is explained why the present research omits the volitional phase of analytical decision-making.

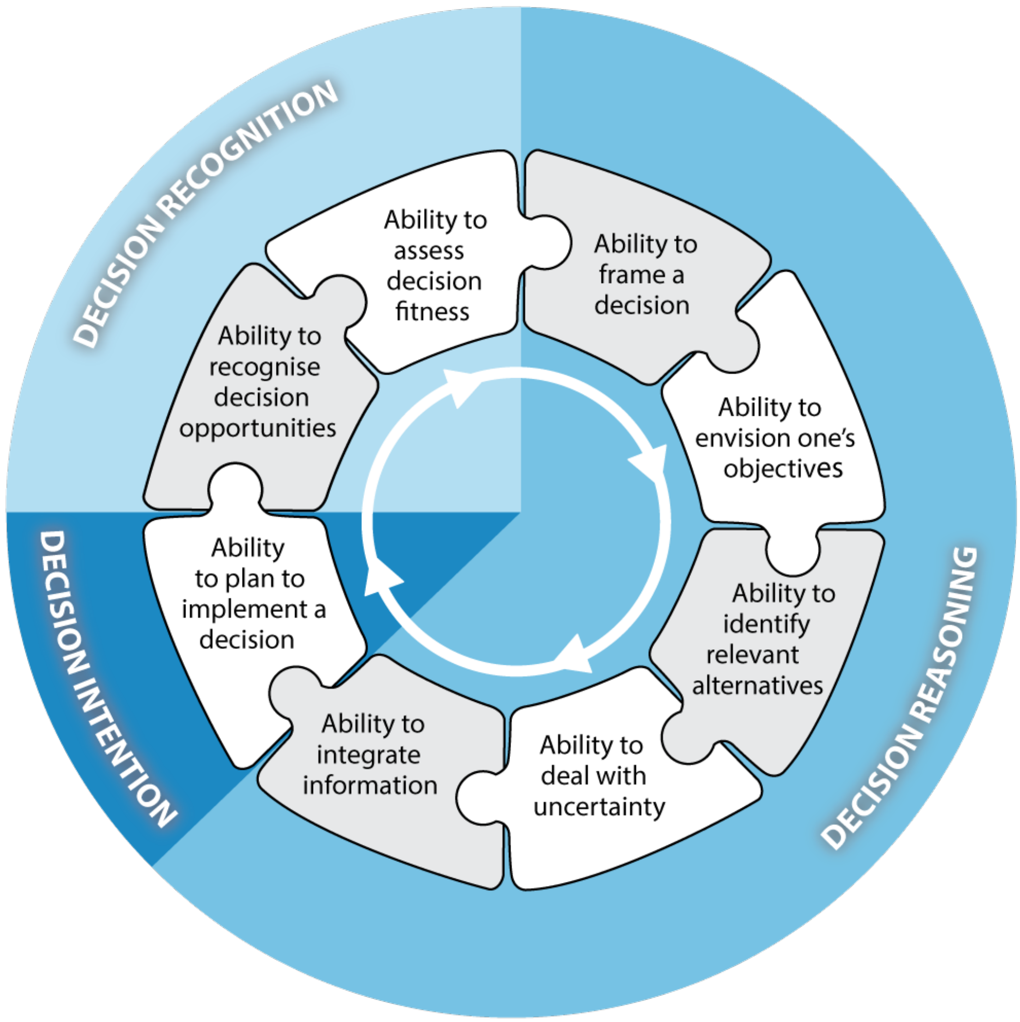


Figure 2-5 Process Cycle of Decision-Analytical Competence

To summarise, the research framework *Process Cycle of Decision-Analytical Competence* in Figure 2-5 focuses on cognitive motivational decision-analytical dimensions of a decision-making process. The framework presents the insights of the prevalent decision analysis literature. The dimensions of the *Process Cycle of Decision-Analytical Competence* define the construct of DAC and have to be examined empirically by this Ph.D. research (cf. chapter 5). The following sub-sections 2.4.1 - 2.4.3 give a theoretical overview of the three process phases and their corresponding dimensions.

2.4.1 *Decision Recognition*

According to the first phase of an individual decision-making process, the *decision recognition*, good decision-makers have to be able to *recognise decision* *opportunities*[27]. Keeney (1992) explains: “instead of sitting and waiting [which could be an explicit choice], it may be preferable to identify decision opportunities, that is, opportunities to better achieve our overall values by formulating a decision situation.” (p. 27) Keelin, Schoemaker, and Spetzler (2009)[28] explain that decision-makers have to declare a decision and be aware of the possibility to configure their future. They must be able to decide whether a particular decision situation needs more attention and is worth further reflection. In this context, decision-makers can decide to “go ... with the flow” (p. 3) and not to decide, or they can decide to make a conscious and explicit choice and take control and responsibility for their decision. Furthermore, good decision-makers are able to assess the potential impact of a decision on their life - whether they face a significant major decision or a routine minor one: therefore, they are also able to balance the effort they should put into making a decision.

Another very important aspect of the first phase of the *Process Cycle of Decision-Analytical Competence* is the ability to *assess one’s degree of decision fitness*. Decision fitness describes a state in which decision-makers are aware of their emotional status and in control of their emotions (Keelin et al., 2009; Thaler & Sunstein, 2008). This state enables decision-makers to focus on the decision, to proceed analytically with the decision-making process, and to make a conscious choice among a useful set of alternatives. Good decision-makers have the ability to know the extent of their competence or decision fitness. Thus, they are able to realise when they need external help or more resources (Keelin et al., 2009). Edwards (1954) and Raiffa (1968) call this ability metacognition.

2.4.2 *Decision Reasoning*

In the second phase, the *decision reasoning*, good decision-makers show the ability to apply decision-analytical rules correctly to implement a sound decision (e.g. Hammond et al., 1999; Keelin et al., 2009). The research field of decision analysis devotes itself to this topic, including belief assessment, value assessment, and integration of beliefs and values (W. Edwards, 1954; Finucane & Gullion, 2010; Raiffa, 1968b). In its second phase, the *Process Cycle of Decision-Analytical Competence* covers five decision-analytical dimensions of good individual decision-making.

---

[27] In comparison to Keeney’s (cf. 1992) proactive definition of decision opportunities, in the present research the term decision opportunities refers more to the recognition of a concrete decision situation rather than creating a situation which enables a decision.

[28] Peer reviewed literature on decision-analytical dimensions of DMC is rare. Even though the work of Keelin et al. (2009) is not peer reviewed and lacks empirical evidence, it is a good theoretical source for this research. The insights of the authors stem from years of practical experiences in teaching young people good decision-making.

First, good decision-makers have to be able *to frame a decision* (e.g. Howard, 2007; Mellers & Locke, 2007). To grasp and understand a decision it is crucial "to answer difficult questions about what to include and, more difficult, what to exclude" (Hastie & Dawes, 2010, p. 31) by drawing a summarising decision tree chart. While considering the context of the decision situation, decision-makers define their decision problem clearly and identify other persons who should get involved as they are somehow affected by the decision and/or its potential output/outcome (e.g. Bazerman & Moore, 2009; Hammond et al., 1999). In addition good decision-makers have to comprehend the decision problem accurately, taking into account what they hope to accomplish by a decision, what is important for making the decision, and which other perspectives of the decision problem they have to consider (Keelin et al., 2009). In this context a good decision-maker has to be able to distinguish between different components of a decision such as uncertainties, alternatives, and (possibly conflicting) objectives.

Second, good decision-makers have to be able to *envision their own objectives* (e.g. Edwards, 1954; Hastie, 2001; Keeney, 1996; Larrick, 2004; Mellers & Locke, 2007; Raiffa, 1968) because these define why the decision-maker is concerned about that particular decision. If decision-makers do not have preferences for what will happen resulting from the decision alternative they choose, they would be satisfied with any decision they make (e.g. Howard, 2000). Hence, decision-makers have to be aware of their individual preferences (Howard, 2007). The more complex the decision, the more carefully decision-makers have to ponder over their objectives. To consider a good list of objectives decision-makers should be creative and intensively thinking at the same time (Keeney, 2007). They have to list and prioritise their short-, mid-, and long-term values (Keelin et al., 2009), articulate what they want and select the most relevant objectives to achieve their goals (Hammond et al., 1999). The literature presents various ways to support decision-makers from generating a comprehensive set of objectives; from simple *wish lists* (e.g. Keeney, 1992, p. 57), taking another person's perspective to think about the decision and creating potential objectives (Keller & Ho, 1988), to structuring one's objectives in terms of separating *fundamental* from *means objectives* (Keeney, 1996). Thus, objectives build an important basis to create alternatives systematically, to guide the methodical search for information, and to plan one's decisions. Using objectives to identify relevant alternatives enhances the likelihood that the decision-maker effectively achieves their desired objectives (e.g. Gollwitzer & Brandstätter, 1997; Gollwitzer & Oettingen, 1998). Furthermore, good decision-makers have to be able to make value trade-offs (Keeney, 1992), which means finding the right relation of costs, such as time, effort and money, to benefits (e.g. Spetzler & Staël von Holstein, 1975). This implies weighting the criteria according to their relative importance in relation to the present alternatives (e.g. Bazerman & Moore, 2009).

Third, good decision-makers are able to *identify relevant alternatives* (e.g. Goodwin & Wright, 2014; Keller & Ho, 1988; Larrick, 2004), i.e. the capability to create and list sound alternatives (Keelin et al., 2009) and to consider almost all desirable alternatives (Hammond et al., 1999) is needed. In this context it is important that the created alternatives fit one's objectives (Keeney, 1992). Several studies have shown that a better and more comprehensive list of alternatives is generated when taking objectives into account to systematically identify alternatives (e.g. Butler & Scherer, 1997; Jungermann, von Ulardt, & Hausmann, 1983; Siebert & Keeney, 2015). Several ways to generate a good list of alternatives are recommended by the literature, for instance thinking about potential win-win situations for decision-involved people and the decision-maker. By trying to achieve one's desires and simultaneously improve the other's situation, thinking of different values could lead to more creativity (Keeney, 1996). Another approach to create a sound list of alternatives is to identify mechanisms that help reach the fundamental objectives of the decision and formulate an alternative for each mechanism (Siebert & Keeney, 2015).

Fourth, good decision-makers have to be able to *deal with uncertainty* (e.g. Edwards, 1954; Matheson & Howard, 1968; Raiffa, 1968). This skill includes the correct comprehension, interpretation, and calculation of probabilities presented in various ways (e.g. G. Gigerenzer, 1996; Grisso & Tomkins, 1996). Making sound decisions implies judging the likelihood of different outcomes and assessing their possible impact (Hammond et al., 1999). Good decision-makers are able to think distributionally and probabilistically by taking an outsider's perspective and perceive a problem not as a one-of-a-kind situation, but rather as a one-case-out-of-hundreds of comparable cases and base their judgements on systematically collected data (e.g. Hastie & Dawes, 2010; Kahneman & Lovallo, 1993). Drawing decision tree charts helps decision-makers to "calculate the decision that leads to the highest expected outcome by applying a rule that follows from decision theory" (Hastie & Dawes, 2010, p. 31).

Fifth, good decision-makers have to be able *to integrate information* within the decision-making process. This implies the ability to consciously consider information needs and gather useful information from the past, present or future to anticipate the consequences of alternatives. "Useful information should come from a credible and unbiased source, be timely, and acknowledge uncertainty. Information about uncertainty ... should recognize the upside and downside risks and their associated probabilities." (Keelin et al., 2009, p. 9) This dimension of a decision-making process also includes the ability to combine assessment of uncertainties and values coherently (e.g. Edwards, 1954; Raiffa, 1968) and make various alternatives and objectives comparable, i.e. "assess[ing] the potential consequences on each of the identified criteria of selecting each of the alternative solutions" (Bazerman & Moore, 2009, p. 3). "For each alternative, [decision-makers have to be able to] take a weighted average of the values assigned to that alternative" (Goodwin & Wright, 2014, p. 34). Therefore, different scales assessing how well the alternatives fit the generated objectives have to be transformed to ease the comparison of assessed alternatives.

2.4.3 *Decision Intention*

Without the *decision intention,* all prior steps are useless (Mellers & Locke, 2007): In the third phase of the *Process Cycle of Decision-Analytical Competence,* good decision-makers *plan to implement a decision* in order to transfer a decision into action (e.g. Hammond et al., 1999; Howard, 2007; Keelin et al., 2009). Decision-makers have to commit to follow through with their decision (Keelin et al., 2009) as it is "[t]he purpose of decision analysis to achieve clarity of action" (Howard, 2007, p. 43). In this respect, coming to a conclusion of which option to choose does not complete the decision process. A commitment to action has to be made and the choice has to be actively implemented. Decision-makers do this by initiating the first steps of action, thereby executing a decision purposefully. To successfully follow through, decision-makers need "resources such as time, effort, money, or help from others. It also requires being prepared to overcome obstacles" (Keelin et al., 2009, p. 11). Within this phase of the decision-making process the motivational and volitional phases of decision-making merge. According to Heckhausen and Gollwitzer's Rubicon model (1987), in moving from the motivational to the volitional phase decision-makers have to exceed the imaginary Rubicon[29] and transfer their intention into action[30]. Even though in the literature intention is seen as one of the main predictors for behaviour, it often explains just 20% to 30% of the variance in behaviour (e.g. Sheeran, 2002). *Intention-behaviour-gap* is what this discrepancy between intention and behaviour is called. It can be observed in people who

---

[29]The model is based on the metaphor of the Robicon (ital. Rubicone), which is a small river in Italy. When 49 B.C. Julius Caesar finally decided to cross the river, the decision for civil war was made and the point of no return was reached.

[30] As described, the ability to plan to implement a decision builds the transition from motivational to the volitional phase (Heckhausen & Gollwitzer, 1987). So there is a second non-cognitive part to this dimension of a decision-making process, which can be simply called action (e.g. Keelin et al., 2009). This step is not captured by these theoretical remarks, not represented in the Process Cycle of Decision-Analytical Competence, and also not an object of the present Ph.D. research.

have positive intentions but yet do not act in that way, even though they show comparable motivation to people who act (Orbell & Sheeran, 1998). Examined by Leventhal, Singer, and Jones (1965), the concept of *action planning* (original term of Leventhal and colleagues) gained publicity from Gollwitzer in the 1990s. Action planning describes common when-where-how plans. It cognitively links a specific behaviour with a specific situation. Thereby, behaviour is connected to situation-specific clues and gets activated almost automatically once the clue(s) appear(s) (Gollwitzer, 1999). As a consequence people with good *action planning* ability act faster (Gollwitzer & Sheeran, 2006) and more often according to their intended aim (Gollwitzer & Brandstätter, 1997). Another concept that seems to play an important role in terms of planning to implement a decision is *coping planning* (Sniehotta, Scholz, & Schwarzer, 2006; Sniehotta, Schwarzer, Scholz, & Schüz, 2005), when a decision-maker should be able to anticipate potential difficulties during the implementation process and has to be able to precisely plan how to overcome those obstacles. Studies have demonstrated that people show more activity according to their intentions if they had made concrete coping plans in advance (e.g. Sniehotta et al., 2006; Ziegelmann, Lippke, & Schwarzer, 2006).

2.5 THEORETICAL COMPLIANCE ON DECISION-ANALYTICAL COMPETENCE

Even if decision analysis does not exhibit empirical evidence for the individual on all theoretical dimensions of DAC so far (cf. sections 2.4.1 - 2.4.3), some research groups have theoretically analysed the construct of DAC, describing the necessary steps of a decision-making process and thereby showing a theoretical agreement on the dimensions of DAC. In the following, four common decision-analytical approaches are representatively outlined:

On the basis of Janis' and Mann's *conflict theory of decision-making* and the corresponding seven criteria of a vigilant decision-maker (1977; cf. section 2.2) Mann, Harmoni, and Power, (1988a, 1988b) developed the *GOFER* process for a decision-making course for adolescents. It was their purpose to simplify the criteria for a good decision-making process and translate them into understandable and acceptable concepts for high schoolers. In this context, GOFER is an acronym for *goals* - envision objectives and values; *options* - taking into account a wide range of alternatives; *facts* - searching for information and collecting data; *effects* – “weighting the negative and positive consequences” of options (Baron & Brown, 1991, p. 64); and *review* – planning to implement the made decision. Later on a sixth step of good decision-making - putting a decision into action - was added to the process.

About 10 years after the GOFER approach, in 1999 Hammond, Keeney, and Raiffa published their book *Smart Choices,* which introduces the decision-analytical approach of *PrOACT*[31] to decision laymen. Corresponding to their approach, good decision-makers have to precisely define their values, identify relevant alternatives, create new options and determine the consequences of each alternative by considering the advantages and disadvantages of each consequence. In this sense, good decision-makers are able to make trade-offs. They clearly evaluate the consequences and how these match the objectives. *PrOACT* also requests the decision-maker to assess uncertainties, identify one's risk tolerance and think about linked decisions.

Similar to the approach of Hammond et al., Howard (2000) explains that the basis of each decision contains three elements: alternatives to decide among, information about what is likely to occur, and individual preferences on resulting outcomes. In addition, the decision-maker has to choose the appropriate frame (e.g. involved people, time point, location, etc.) for their decision. This frame influences every element of the decision. The elements of a decision are presented by the *six elements of a quality decision* (Keelin & Spetzler, 1992). Keelin and Spetzler's model of decision elements contains the following dimensions: helpful frame, clear values, creative alternatives, useful information, sound reasoning, and commitment to follow through.

In comparison to the *six elements of a quality decision,* the *Socio-Technical Decision Analysis* of Phillips and colleagues (e.g. Phillips & Bana e Costa, 2007; Phillips & Phillips, 1993; Phillips, 2007) serves first of all group decisions, but is also applicable to individual decisions. Comparable to the presented approaches of Hammond et al. (1999) and Keelin et al. (2009), at the beginning of Socio-Technical Decision Analysis decision-makers have to recognise a need for decision-making and the decision problem has to be defined and structured. Subsequently, all existing and potential alternatives and criteria, which seem to be relevant or attractive to the decision-maker, have to be collected and listed. By building a criteria-tree, the decision criteria are analysed and the options are scored. The process of judging the weights of the criteria across and within each other is called *swing weighting*. The main aim is to estimate the importance of the criteria in relation to each other. The final data analysis also includes a sensitivity analysis.

It is important to remark that the approaches from different research groups offer a similar process with similar elements despite using different terms for these elements.

[31] The acronym stands for problem, objectives, alternatives, consequences and trade-offs, which represent the steps of a rational decision-making process.

Mellers and Locke (2007) emphasise this impression in their literature review of decision analysis by identifying five interrelated steps suggested by decision analysts to make good decisions: "[d]efine the problem and set the goals, gather information and identify options, evaluate the information and the options, make a choice and implement the choice and monitor the results." (p. 351-352) In general, all approaches define the decision problem, determine objectives, create alternatives, and gather information. In this sense, a common theoretical signification of what is meant by a sound individual decision making process is available.

To summarise, more than 50 years of research in the field of decision analysis has created a well-established basis of decision-making insights and understanding. Different approaches of various research groups from decision analysis provide similar steps for a good individual decision-making process. Thereby, the different research groups show a shared knowledge and partial agreement of what a sound analytical decision-making process of an individual should look like and build the theoretical basis for the presented Ph.D. research. From the existing literature in decision analysis, the term DAC, used in the present research, refers to the association of eight motivational cognitive abilities. The following sub-section is intended to provide an overview of the eight cognitive dimensions of the construct DAC.

2.6 The Dimensions of Decision-Analytical Competence

On the theoretical basis described above, the construct of DAC is defined with its eight cognitive motivational abilities listed in Table 2-2.

Table 2-2 The eight cognitive dimensions of DAC

Phase of the Process Cycle of DAC	#	Cognitive DAC-dimension
Decision recognition	1	Ability to recognise decision opportunities
	2	Ability to assess decision fitness
Decision reasoning	3	Ability to frame a decision
	4	Ability to envision one's objectives
	5	Ability to identify relevant alternatives
	6	Ability to deal with uncertainty
	7	Ability to integrate information
Decision intention	8	Ability to plan to implement a decision

Note. DAC = Decision-analytical competence.

Consolidated in the phase of *decision recognition,* the ability to *recognise decision opportunities* and the ability to assess decision fitness are displayed, followed by the phase of *decision reasoning* consisting of the abilities to *frame a decision*, to *envision one's objectives*, *identify relevant alternatives*, *deal with uncertainty* and *integrate information*. The eight skills are completed by the ability to *plan to implement a decision* in the phase of *decision intention*. Based on this definition, the present research aims to examine how the construct of DAC can be operationalized and measured.

The next chapter explains the chosen research methods for designing a psychometric performance test and presents the research hypotheses. By introducing a strategy for data gathering and the corresponding target group, it illustrates how empirical data was collected.

3 RESEARCH METHODOLOGY

As mentioned before, the main target of the present research is to provide a reliable and valid psychometric performance test, which captures the construct of DAC with its cognitive motivational dimensions (cf. Table 2-2). After theoretically defining the DAC-dimensions (cf. section 2.4) a psychological standard procedure was used to develop the test (cf. section 4.1). Before presenting the operationalization of DAC, the subsequent sections serve to give a compact overview of the chosen research method for creating a psychometric performance test. Thus, the general purpose of this chapter is to introduce the research hypotheses, the research methodology and its corresponding methodological decisions, as well as the strategy of data collection and the target group (cf. section 3.3).

In psychology a *psychometric test,* is "any standardized instrument, including scales and selfreport inventories, used to measure behavior or ... cognitive abilities (reasoning, comprehension, abstraction, etc.), ... and personality characteristics." (Zedec, 2014, p. 278) While the items of a test represent the *independent variable(s)*[32]*,* the test response set suggests the *dependent variable*[33]. Figure 3-1 gives an example and shows a cutting of the non-verbal multiple-choice measure called *Raven Progressive Matrices test* (Raven, Raven, & Court, 1962), which aims to capture *analytical intelligence*, that "refers to the ability to deal with novelty, to adapt one's thinking to a new cognitive problem" (Carpenter, Just, & Shell, 1990, p. 2): An indicator represented by *item 1* constitutes the independent variable. The response options (*a – h*) represent the dependent variable. To answer the item, the test participant has to choose between eight options - the response set. According to the test construction, selecting option *b* would be interpreted as high occurrence of the dependent variable – a high parameter for analytical intelligence – as *b* is the correct answer offering the correct requested geometric form in combination with the fitting line texture and orientation. Selecting any of the other options would lead to the assumption of a low occurrence of the dependent variable – a low parameter for analytical intelligence – as those options are incorrect.

[32] Independent variables represent inputs or causes of a phenomenon.

[33] Dependent variables represent outputs or effects of a phenomenon.

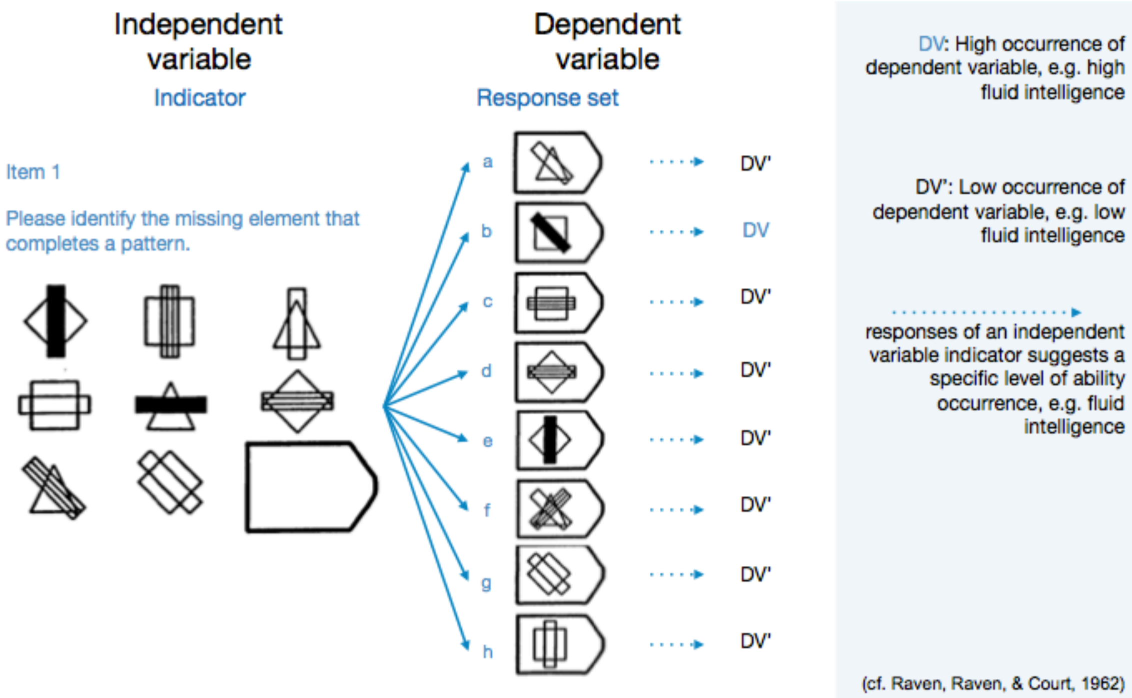


Figure 3-1 Independent and dependent variable exemplified by an item of a performance test

The *classical test theory* (cf. for instance Guliksen, 1950; Novick, 1966) assumes that a person's behaviour can be explained by individual traits or characteristics (cf. *psychological constructs*; Cronbach & Meehl, 1955). Consequently, those traits or characteristics can be measured by a test, which consists of several items or so-called indicators (e.g. Moosbrugger & Kelava, 2012). In the case of the present research objective, indicators are chosen as they rely on *test data* (cf. Cattell, 1965). If the scientific literature gives reason to assume that a latent construct encompasses several dimensions, a corresponding psychometric test should capture all of these dimensions.

In the framework of the present research it is assumed that DAC is a higher-order factor, which underlies various dimensions (Law & Mobley, 1998) as presented in section 2.4. All cognitive DAC-dimensions build the dependent variables, which are operationalized by a set of items/indicators. Before the operationalization of each dimension is introduced in section 4.1, the research hypotheses (cf. section 3.1), the process of psychometric test development (cf. section 3.2), and the strategy for data collection (cf. section 3.3) are introduced in the following sections and sub-sections.

3.1 RESEARCH HYPOTHESES

By developing a set of decision tasks that capture decision-analytical approaches to deal with complex decision-making, which are widely recognised in the corresponding literature, the main aim of this research is to successfully operationalize and validate a performance test for DAC. Whether these expectations are met and thereby whether it can be claimed that DAC can be measured by a set of decision-analytical tasks, is examined by the following hypotheses (cf. Table 3-1). Hypotheses H_1 – H_3 are statistical-methodical hypotheses and hypothesis H_4 is a content-valid hypothesis, which is further specified in sub-sections 5.1.1 to 5.1.7 of this thesis.

Table 3-1 Research hypotheses

Type	#	Hypothesis	Comment
Statistical-methodical	H_1	Each dimension of the DAC-test is reliable and thus shows internal consistency.	Internal consistency is a measure based on the correlation between various items of a homogenous test (capturing only one dimension, cf. section 3.1.1) or items of a sub-scale of a test of a higher-level measure (e.g. Lienert & Raatz, 1998, p. 192). It shows whether different items produce similar scores.
	H_2	The DAC-test is reliable and shows internal consistency.	Reliability can be defined as the degree of consistency with which a test measures a specific trait, independently of whether the test aims to measure this specific trait or not (e.g. Zedec, 2014, p. 307). There are different approaches to measure the reliability of a test. In the context of this research, the internal consistency of the constructed DAC-test is investigated by Cronbach's alpha.
	H_3	The DAC-dimensions refer to one general factor.	According to the literature, the various DAC-dimensions reflect one common concept of decision-making. The current research intends to examine this assumption by both an exploratory and a confirmatory factor analysis. The statistical procedure of an *exploratory factor analysis* describes the variability among observed correlated variables in terms of a potentially smaller number of unobserved latent variables (Bortz & Döring, 2005, p. 355). These latent variables are called factors. Confirmatory factor analysis describes "any method of testing a priori hypotheses to the effect that the relationships among a set of observed variables are due to a particular set of unobserved variables." (Zedec, 2014, p. 58).
Content-valid	H_4	The DAC-test score correlates significantly with suitable criteria for validation.	Validity can be defined as the degree of accuracy with which the test measures the specific construct it is supposed to measure (Lienert & Raatz, 1998, p. 10). To prove the validity of a test, the correlation of the test with suitable criteria for validation has to be examined, such as decision self-esteem or vigilant decision-making style.

Note. H = Research hypothesis; DAC = Decision-analytical competence.

Whether the hypotheses are corroborated or have to be rejected, is discussed in section 6.2.

3.2 PSYCHOMETRIC TEST DEVELOPMENT

For the following compact overview[34] of a standard procedure of psychometric test development, the present information, explanations, and statements delve into the area of psychometric test design.

According to Lienert and Raatz (1998), the psychometric procedure of test development can be divided into six phases as displayed in Table 3-2, which serve to structure the following sub-sections. The first phase is called *test design* (cf. sub-section 3.2.1), in which the test concept and test form are chosen, and the scope of application and the target group are defined. Additionally, a decision about the homogeneity or heterogeneity of a test is made. In the phase of *item development* (cf. sub-section 3.2.2) decisions on the item types and response sets have to be made, the items have to be phrased, and the item goodness criteria have to be considered.

Table 3-2 Standard procedure of psychometric test development (cf. Lienert & Raatz, 1998)

#	Phase of test development	Elements of test development phase	Sub-section
1	Test design	Generating the test concept, deciding about test homogeneity or heterogeneity, defining the scope of application, choosing the test form and the target group	3.2.1
2	Item development	Choosing item types and response sets, item phrasing, considering item quality criteria	3.2.2
3	Item analysis	Pretesting the items	3.2.3
4	Distributional analysis	Selecting items, revising items, calculating the test score, analysing the test distribution, standardisation, developing the final test	
5	Empirical control	Analysing the reliability, analysing the validity	3.2.4.
6	Test calibration	Interpreting the test results, defining norm values, test calibration	

The pre-testing of items in a first test trial is called *item analysis*, followed by the *distributional analysis* (cf. subsection 3.2.3), in which the best items are selected by their quality criteria. Where necessary items are revised, the test score is calculated, and the test distribution is analysed. After the standardisation, the final test has to be developed.

[34] As scientists have published books on the topic of psychometric test development, the present chapter gives just a compact overview of the most important methodological steps/decisions to provide the reader with the necessary information to understand the thoughts and considerations on the chosen research method.

In the phase of *empirical control,* the test's reliability and validity are analysed. The test result interpretation, definition of norm values, and the calibration are called *test calibration*. Standardisation and test calibration are presented in sub-section 3.2.4.

For developing the DAC-test, the first five[35] of these six phases were executed. Several decisions on methodical issues were included within this process. The following sub-sections explain the development of the DAC-test and justify the choices made.

3.2.1 Test Design

At the beginning of any research project, the object of research has to be defined. For the present research this step is executed in chapter 2, in which the research domain is presented by the literature review and the object of research is captured – here DAC.

The main aim of the test construction strategy for this research was to capture the competence of analytical individual decision-making and thus to create a pool of content valid items. To pursue this aim, a combination of *rational/deductive* and *criterion oriented/external test concept* (e.g. Bühner, pp. 93-94, 2010; Eid & Schmidt, 2014, pp. 57-60) was chosen. Therefore, the decision-analytical literature was searched for either empirical studies that examine group differences in decision-making (*criterion oriented/external* approach), or for well-founded theories about the to be measured construct – DAC (*rational/deductive* approach). If prevalent studies did not provide a full set of appropriate items to operationalize a cognitive dimension of DAC as, for example, the work of Bond et al. (2008, 2010) does for the dimension *ability to envision one's objectives,* new items had to be generated.

According to the theoretical concept of DAC, it is seen as a multidimensional construct that underlies eight dimensions (cf. Table 2-2). Parker and Fischhoff argue "In principle, these decision-making skills could be independet of one other. However, the cluster identified in previous studies of individual differences overlap one another (2005, p. 8). Consequently, a corresponding psychometric performance test had to be *heterogeneous*[36] to capture all of its dimensions and enable a broader *scope of*

---

[35] The present research provides a first attempt to psychometrically assess DAC. As the corresponding empirical results have to be treated more as descriptive than as normative, it is not intended to execute the final phase of psychological test development, the so-called calibration, which would demand a pre-defined standard of comparison.

[36] The tighter the scope of application, the more homogeneous the test. The broader the scope of application, the more heterogeneous the test (Lienert & Raatz, 1998).

application. The scope of application of the DAC-test relies on a specific decision type that individuals are facing. The characteristics of this decision type refer to the kind of decisions, for which decision analysis can be instrumental. In this sense, based on Matheson and Howard (1968, p. 3), the DAC-test is supposed to measure the ability to make decisions that show the following characteristics:

- Complex preferences: The desires of the decision-maker are not clearly formulated and might be conflicting.
- Costly: Many resources (time, effort, money) are involved in the decision process.
- General: The decision situation refers to decisions normally appearing in every person's life, not only to a specific type or topic of decision.
- Important: The life of the individual will be affected by the results of the decision for many years.
- Personal: The decision-maker makes a private decision for himself or herself, not on behalf of an organisation or another person.
- Rare: The decision problem appears merely a few times or even only once in a lifetime.
- Uncertain: Many key factors of the decision are imperfectly known by the decision-maker.

To operationalize DAC, a *performance test* as a measuring procedure was chosen. As explained in section 2.2, a performance test measures the actual performance of a specific motor, sensorial, or cognitive ability in contrast to a questionnaire, which contains self-reported opinions, attitudes, and interests and/or assesses a person's personality traits (e.g. Zimbardo & Gerrig, 1996). Performance testing addresses several challenges that questionnaires/self-assessments raise:

In a self-assessment the subjects can influence the questionnaire results in their favour. Such manipulation can happen in two ways: due to *impression management* – a conscious manipulation of the test results in order to draw a positive picture of oneself (e.g. Goffman, 1956), or due to *self-deception* – a "positively biased response that the respondent actually believes to be true" (Paulhus, 1986, p. 144). A common and well-examined example for the first way is the *social desirability bias* (e.g. A. L. Edwards, 1982; Ellis, 1946). In this case, test participants answer items in a manner that will be seen as beneficial by others. It appears in the form of underreporting or overemphasising undesirable traits or concrete behaviour. When self-deceiving, examinees' answers in a questionnaire are guided by underlying unconscious and situational consistent self-images (e.g. Cheek & Hogan, 1983). “Although organized for a positive self-presentation, these images do not involve conscious dissimulation” (Paulhus, 1986, p. 144).

Influencing a performance test in either way is only possible in terms of underperforming. People cannot show better skills than they actually have[37]. In addition to the challenge of impression management and self-deception, self-reports run a higher risk of becoming unreliable and invalid due to lacking insights or sensitivity to the demanded topic (e.g. Finucane & Lees, 2005; Nisbett & Wilson, 1977). In contrast, self-reports are often not as strenuous for the examinee as they usually require “first thing that comes up your mind” answers in comparison to performance tests’ mostly postulated “well-thought” responses. However, for this research a performance test was chosen, also since a performance test showing good statistical properties is absent in this research area.

Additionally, it was decided that the test should be provided as an online test. Besides its obvious advantage of allowing “access to much wider populations ... in an inexpensive, fast, and convenient way” (Dandurand, Shultz, & Onishi, 2008, p. 428), online testing can also be done in diverse settings – location- and time-wise (Reips, 2001, 2002a) and thus, it enhances the participation comfort for examinees (Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003). So far, only a few studies have been undertaken to compare online and paper-pencil versions of ability tests. In this respect, the research of Preckel and Thiemann has verified that “valid and reliable data can be gained through online ability assessment” (2003, p. 137). The disadvantage of risking multiple submissions (Reips, 2001) was antagonised by using IP address verification (Reips, 2002b) in the present study. Monetary incentives (Bosnjak & Tuten, 2003) and direct feedback were applied to counteract potentially high dropout rates. Furthermore, in contrast to paper-pencil test, studies have shown that online tests provide fewer missing data (Stanton, 1998) and they are not that susceptible to socially desirable responding (Rietz & Wahl, 1999).

Young adults from the age of 18 to 30 years were selected as a target group for this Ph.D. research. The criteria for choosing this group are presented in sub-section 3.3.1.

[37] Thus, for instance, the researcher can ask a small man how tall he is and probably gets an answer in which the man adjusts a little upwards, since in society men are supposed to be taller than women. The researcher can also measure/test his body height. If the man does not trick the researcher by standing on his tiptoes, the result cannot be influenced by the test participant.

After generating the test concept, deciding about test homogeneity and heterogeneity, defining the scope of application, and choosing the test form and the target group, the items had to be developed. The following sub-section presents the various considerations and decisions for the item development. The actual item generation is presented in section 4.1.

3.2.2 Item Development

Following the phase of *test design,* the items for the DAC-test had to be developed. According to the theoretical concept of DAC (cf. section 2.4), the corresponding test had to be heterogeneous to capture each cognitive dimension. In comparison to existing tests such as the A-DMC (Bruine de Bruin et al., 2007) or the Y-DMC (Parker & Fischhoff, 2005), which with regard to content contain disconnected tasks (cf. Figure 2-3), the DAC-test presents a set of content-wise connected items (cf. section 4.1). The underlying aim of this decision was to approach the representation of a complex real world decision situation, which requires the performance of all cognitive abilities, i.e. DAC-dimensions, as closely as possible.

Figure 3-2 presents a schematic picture of the DAC-test's intended structure. As illustrated here, it was intended that each block of tasks demands only the performance of one DAC-dimension at a time, and therefore allows for analysing each dimension independent of the other dimensions. For those cases in which the decision science literature provides an approach for measuring a DAC-dimension that conforms to the theoretical assumptions underlying this research, the existing approach was used to operationalize this particular DAC-dimension (cf. sub-section 4.1.1). If prevalent studies did not provide a theory appropriate set of items to operationalize a DAC-dimension, new items had to be generated.

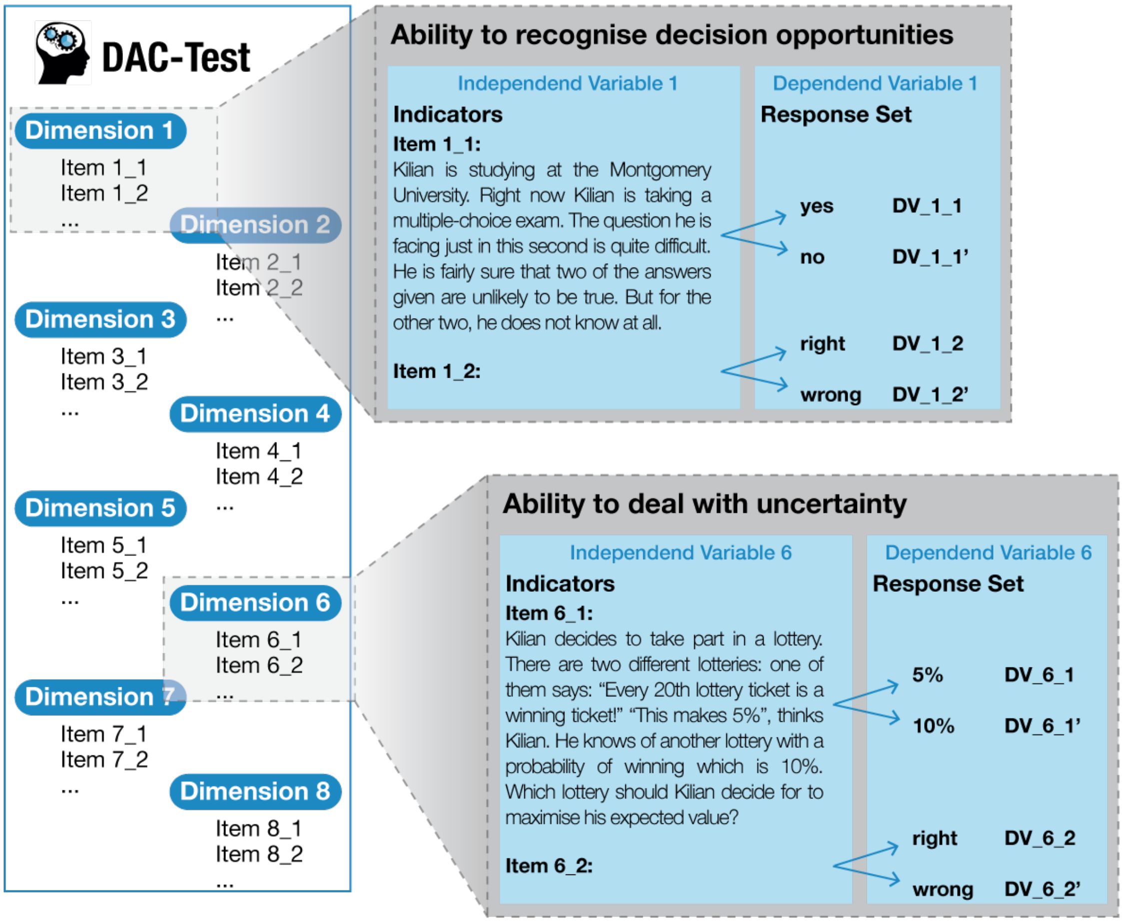


Figure 3-2 Schematic picture of the structure of the DAC-test

Figure 3-2 shows on the left the test with its eight cognitive dimensions, of which each consists of a set of indicators, i.e. items. One of the most important aspects of test development is the conjunction between its items and the underlying theoretical construct(s) (Porst, 2011). To quantitatively and qualitatively fit the purpose of research, items were chosen to content-adequately operationalize and exhaustively capture the construct of DAC. In terms of the *item type(s)* and *response set(s),* items with bounded responses and single referencing, i.e. only one possible correct answer at a time, were chosen, which are characterised by a limited, predefined and exhaustive response set. For each response set *distractors* had to be defined, which describe the incorrect answer options of response sets that aim to build potential alternatives to the correct answer. "[I]n item trials, ideally, each distractor should be equally used by subjects failing the item. Obviously, as distractors in the options become useless, so an item becomes easier and easier." (Kline, 2015, p. 36) The great advantage of *bounded items* is their efficiency and objectivity in the phase of questionnaire completion as well as in the phase of data handling and analysis (e.g. Kline, 2015; Mossbrugger & Kelava, 2012). Its disadvantage lies in running the risk of non-responses for some items as examinees cannot fit their answer into the response set and therefore skip the answer or just select randomly. To minimise this effect, responding to each item was forced, i.e. it was impossible for the examinees to proceed with the test without answering the missing item. Additionally, some selected items used *formula scoring* in comparison to *number right scoring*, which "attempts to reduce the influence of random guessing on the test score. With this scoring method a 'don't know' answering option was added to the true-false items, and the number of correct minus incorrect answers ... is used as the test score." (Muijtjens, van Mameren, Hoogenboom, Evers, & van der Vleuten, 1999, p. 267). *Formula scoring* advises test participants to skip an item if they are persuaded that their response would be equal to a haphazard guess, as true-false items have a 50% probability of choosing the correct answer just by guessing (e.g. Kline, 2015). In comparison to number-right scoring formula scoring has higher test reliability (Muijtjens et al., 1999).

In terms of item phrasing, special attention was paid to ensure that each item (block) requires just one of the cognitive decision-analytical abilities to be deterministic. Additionally, items were constructed so that they were as short and intelligible as possible, avoided foreign words and addressed critical ethical issues (e.g. Payne, 1951; Schnell, Hill, & Esser, 2008).

Regarding the issue of which specific item scale was chosen (whether *nominal, ordinal, interval and ratio scale level* (Stevens, 1946) for each DAC-dimension, cf. sections 4.1.1 - 4.1.8. As the literature does not claim one scale type being better than any other (e.g. Porst, 2011), each single item type had to be inspected in order to choose the best fitting scale type. Independent of the scale types, care was taken that the response set(s) were clearly laid out, adequate to the complexity of the items, and free of ambiguous quantifiers such as "sometimes" or "often" (Lienert & Raatz, 1998).

A test can only be as good as its items. Consequently, for a well-designed test, the test and its items have to meet *criteria of quality*. In this context, an item can be assumed to be objective, if different examinees with the same extent of the DAC-dimension answer the item in the same way. An item is assumed to be valid if examinees with a strong DAC degree of occurrence answer the item more often according to the expectations of the researcher - "in key direction" - in comparison to examinees with a weak DAC degree of occurrence. A high correlation between an item and the overall score of the test is called *discriminatory power*, i.e. that the test "achieve[s] a good spread of scores" (Kline, 2015, p. 8).

Another item quality criterion considered is called *item difficulty*. Standardised tests, as for instance the Wechsler Adult Intelligence Scale-Revised (Wechsler, 2008), usually have a medium degree of difficulty. Therefore, the test consists of items with low (the majority of examinees can solve the item), medium, and high (only a few examinees can solve the item) difficulty.

Regarding the test and data administration, several aspects were taken into account:

First, the test instruction (cf. chapter 4) was written to be as comprehensible and transparent as possible to the examinees. Accordingly, on the welcoming page of the DAC-test the research object and its importance for research is briefly explained. Information on the researcher is given and details of privacy protection, in which confidentiality is ensured, are presented. Additionally examinees are informed about the advantages of taking the test (cf. section 3.3), and what to expect and which testing aids are allowed are clarified.

Second, for the decision of how many items to choose per DAC-dimension for the first version of the DAC-test two criteria were considered: The duration of the test should be kept as short as possible as concentration and motivation of examinees decrease with increasing time (Wise & Kong, 2005). And, to be able to select the best compatible items for the final DAC-test, a sufficient variety of items should be tested.

Third, the digital data storage was done on a *MySQL*[38] open source database, which was chosen as it allows data export in various common data formats and conforms to the recent status quo of technical standards (Suehring, 2002).

Fourth, in terms of statistical data analysis the free software environment for statistical computing and graphics named *R* was used. The programme contains a variety of possibilities to organise, transfer, or analyse data and generate corresponding graphics. Instead of providing a user interface with specified menu items, as found for instance in the popular statistics software *SPSS* (IBM Corp, 2012), *R* users have to actively produce commands.

Closing the phase of *item development,* a first version of the DAC-test was designed. In order to analyse its items according to their statistical fit, the prototype had to be tested. By selecting the best fitting items, calculating the test score and analysing the test distribution the final DAC-test was developed. Sub-section 3.2.3 presents the procedure of these next two steps of the test design – the so-called *item and distributional analysis.* The statistical analyses and actual results are shown in section 4.2.

[38] MySQL "is the world's most popular source database" (MySQL Editions, n.d.)

3.2.3 Item Analysis & Distributional Analysis

Even though pre-testing belongs to the most common procedures for test improvement, the methodological literature does not provide *one* universal standard procedure (Prüfer & Rexroth, 1996). In the case of the present research, the prototype of the DAC-test was qualitatively pre-tested by eight subjects belonging to the target group before being quantitatively pre-tested, as described below.

Qualitative pre-testing

The main purpose of the *qualitative pre-testing* phase was to verify the intelligibility of item wording, sentence structure and response sets, as well as of the instructions. Therefore, a multi-method pre-testing procedure as suggested by Prüfer and Rexroth (1996) and illustrated in Figure 3-3 was applied.

Figure 3-3 Procedure of the multi-method qualitative pre-testing phase

In the first phase of the qualitative pre-testing, the DAC-prototype was presented to a sample of eight people individually. Each of them took the test at home in the presence of the researcher. The examinees were told that a prototype is presented and that they have a special role helping to improve the test and its procedure. While the examinees were taking the DAC-test online on their own computer, survey protocols about spontaneous questions and comments of the test participants, behavioural observations, technical problems with the test, and the testing time were minuted.

The second phase of qualitative pre-testing was performed immediately afterwards. An evaluation questionnaire (cf. appendix 9.2) was handed to each examinee aiming to gain more information on non-observable problems as well as motivational aspects of the test. According to Kuhnke (2007), the evaluation questionnaire measures the following parameters: degree of interest in the topic; degree of fun during the test completion; personal judgement about the length of the test, degree of understandability of the items, response sets and instructions; degree of difficulty of the various item tasks; occurrence of delicate[39] items; general assessment of the test; and suggestions for improvement.

In the third and last phase of the qualitative pre-testing, qualitative semi-structured interviews (for the interview guideline cf. appendix section 9.3) took place referring to the individual test results, the test protocol, and evaluation questionnaire. In comparison to the evaluation questionnaire, the interview allowed a higher degree of freedom and adaptability in terms of acquiring relevant information from the interviewee (e.g. Bortz & Döring, 2005). The interview also served to clarify comprehension questions.

Based on the analyses of the feedback packages of the eight people - test results, test protocol, evaluation questionnaire, and interview – it was decided which items to adapt in terms of their content, wording, phrasing, response set and/or layout, substitute or cancel, and which items to keep in their initial form. Additionally, the test instruction was revised and minor technical problems with the online test were solved. At the end of the qualitative pre-testing, the first DAC-test prototype was blue-pencilled in order to start the quantitative pre-testing and the actual item analysis with the second DAC- test prototype.

Quantitative pre-testing

As the item analysis serves as a test trial, the *quantitative pre-testing* was operated under the same conditions as planned for the main study. Aiming to select those "items that form a homogeneous, discriminating scale" (Kline, 2015, p. 133), the second DAC-test prototype was presented to a total sample of 196 subjects[40] representative of the target group. Therefore, the correlation of each item with the total score of its scale and the proportion of the sample group answering according to keyed response was calculated. This step of the pre-testing – the actual *item analysis* - serves as descriptive-statistic evaluation of the prototype. Its main aim is to increase the test's reliability and validity by revising test items and proving the test point distribution.

[39] The adjective delicate describes items that cover either ethically controversial topics as for instance abortion, or topics that are in general perceived as being very personal such as for instance sexual preferences.

[40] According to Lienert and Raatz (1998), an appropriate sample size for the item analysis would be 200-400 subjects. According to the dimensions of Ph.D. research – time- and money-wise - the lower boundary seemed to be more realistic to achieve.

Two essential criteria for assessing the item quality are the *discriminatory power coefficient* (r_{it})[41] that describes the correlational relation between the item scores and the test scores, and the item *difficulty index* (P_i)[42] that defines the percentage frequency with which an item is answered correctly by the group of examinees (e.g. Kline, 2015; Lienert & Raatz, 1998). On the basis of these two essential statistics and according to Kline (2015) all items meeting both criteria were picked for the final test. Items that failed one criterion were examined for particular characteristics that could have caused this non-fit. Furthermore, it was ensured that all aspects of each DAC-dimension were covered and that all DAC-dimensions provided approximately/not less than eight items, as this enabled good-accessible comparability of scores between examinees.

The item analysis also serves to test the sufficient distribution of test points and thereby whether the construct is normally distributed. As the composition of the sample group of examinees was compatible and the display of the frequency distribution justified the assumption of a normal distribution, it was assumed that the DAC-test scores are normally distributed.

At the end of the item and distributional analysis the final version of the DAC-test was assembled. The next sub-section presents how reliability and validity of the final test were examined in the test construction phases of *empirical control* and *test calibration*.

3.2.4 Empirical Control & Test Calibration

The final version of the DAC-test was presented to a representative sample of several hundred young adults to empirically measure reliability and validity of the test.

"There is virtual consensus among researchers that, for a scale to be valid and possess practical utility, it must be reliable." (R. A. Peterson, 1994, p. 381) Thus, the items of a scale have to be consistent, i.e. the results have to be free of error and produce comparable results. As mentioned in sub-section 3.2.2, reliability can be defined as the degree of accuracy with which a test measures a specific trait, independently of whether

[41] A r_{it} close to 1 (or -1 if the items are negatively poled) means that examinees with a strong trait occurrence solve the test with a high test score x_v, and examinees with a weak trait occurrence solve the test with a low test score x_v. A r_{it} between .4 and .7 describes a "good" discriminatory power. In this case, the items seem to differentiate similarly to the test. A r_{it} close to 0 means that the differentiation by an item does not correlate with the differentiation of the test. The item is not suitable to differentiate between examinees with a strong trait occurrence and examinees with a weak trait occurrence.

[42] The P_i lies between 0 and 1. The more examinee that solve an item the higher the P_i. This means, the higher the P_i the easier the item is to solve: A P_i of 0 means that no examinee solved the item. It is too difficult. A P_i of 1 means that all examinees solved the item. Thus, it is too easy. Ideal would be a p_i of 50 (e.g. Kline, 2015).

the test aims to measure this specific trait. In the framework of the *classical test theory*[43], reliability[44] is defined as the quotient of the variance of the true/ideal score and the variance of the observable score. Given that reliability relates to the consistency and stability of a test's results, potential measurement errors were kept as low as possible (e.g., Homburg and Giering, 1996; Remenyi et al., 1998). Although online testing does not allow for the standardisation of test circumstances, such as the test environment and the time point of measurement, an attempt was made to standardise the test procedure and the test analysis to achieve a high reliability (Lienert & Raatz, 1998). In order to approach standardisation for the test procedure standardised invitations were sent, all necessary information was built in the test, and direct feedback was given right after submitting the test. Concerning the standardisation of the test analysis, one spreadsheet for editing the data was used and the analysis was run with the same previously generated R-script.

Based on the established literature (e.g. Lienert & Raatz, 1998; Moosbrugger & Kelava, 2012), there are various approaches to test for reliability. The present research intended to test internal consistency by *Cronbach's alpha* (Cronbach, 1951) and *parallel-test reliability* in comparison to retest reliability.

Cronbach's alpha is theoretically equivalent to the mean of all possible split-half reliabilities. If a test contains homogeneous items, each item can be seen as part of the test to measure the corresponding trait. The higher the correlation of the items with each other on average, the higher the internal consistency of the test. The advantage of Cronbach's alpha is that neither a parallel test version nor two testing times per examinee are needed. Additionally, the overestimation of reliability due to trainings effects can be prevented. As examinees do not have to take the same test twice, learning or memory effects, which affect the performance of test takers, can be avoided.

As a second indicator for reliability, parallel-test reliability should be measured. Therefore, two equivalent test forms, which are content-wise as equal as possible, have to be developed. Two tests can be seen as so-called *twin tests*, if their item sets show

---

[43] The classical test theory is the most common psychometric test theory. For the present research it was particularly interesting to examine the higher-order structure of the DAC-dimensions. This is not a particular strength of item response theory but of structural equation modelling. The item response theory focuses on estimating item parameters like item difficulty and item discriminatory power while structural equation modelling focuses on factor structure per se. Consequently, factor analyses and classical test theory were applied.

[44] As a theoretical variable, reliability is clearly defined and lies between 0 and 1. Ideally, the reliability of a test is as high as possible. Performing variables, for instance the Intelligence Quotient (Wechsler, 2008; as well as expected for DAC), are easier to measure precisely in comparison to trait variables, such as for instance openness to new experiences. Hence, well-established performance tests reach a reliability of .90 or even .95 in comparison to well-established trait tests which often merely attain a reliability of around .70.

comparable variances of the test scores, i.e. if they result in the same mean values. The parallel-test reliability is calculated by the correlation between the two test forms (e.g. Moosbrugger & Kelava, 2012).

Besides reliability, validity is one of the key goodness criteria for psychometric tests. As mentioned in sub-section 3.2.2, validity can be defined as the degree of accuracy with which the test measures the specific trait/ability it claims to measure (e.g. Cronbach & Meehl, 1955). Different aspects of validity can be distinguished: criterion, content, and construct validity (e.g. Kane, 2001; Kline, 2015; Messick, 1995). In modern validity theory, *construct validity*[45] builds the central aspect overarching all other types of validity evidence (e.g. Messick, 1995).

To empirically measure the construct validity in the framework of the present research, a correlative approach was chosen. Therefore, the correlation of the test score with various manifest variables, such as another test for decision behaviour, is calculated based on a theoretical well-founded assumption about the direction and height of the correlation. If the empirical correlation is suitable for the theoretical interdependences, it can be assumed that the test results can be attributed to the latent construct. Cronbach and Meehl call this a hypothetical-deductive approach. The chosen manifest variables, the so-called validation criteria, with which the test score of the present research were correlated, are presented in sub-sections 5.1.1 - 5.1.6.

According to the content validity of the DAC-model (cf. Figure 2-5), eight theoretical dimensions of the DAC have been addressed by the literature review of the present research. The DAC-dimensions have been selected to capture specific aspects of the analytical avenue to DMC. Even though there are more components, such as making trade-offs, account for risk attitude and time preferences, updating beliefs or inference, the DAC-dimensions have been selected on theoretical and methodological considerations. One of the main intentions of the developed performance test was to construct a coherent case study, which represents a complex decision case close to real life and whose successful solving requires a set of cognitive abilities. It could be assumed that each dimension measures a relatively distinct cognitive process of decision-making.

Additionally, for the testing of the DAC-test's quality criteria, the *standard of comparison* can be defined. When a subject takes a test, a numeric test result is received

[45] Construct validity is defined as the degree of accuracy with which interpretations of test results can be made as defined by explanatory concepts.

– the test score. To classify this individual test score and interpret it in accordance with the existing range of results, a norm value of the to be measured construct has to be determined. The assessment of an individual's test result in relation to the performance of a sample group is called *norm-oriented interpretation.* It is common that at this point of test development no test score distribution of a norm sample is available. Another approach is to interpret the results in relation to ex-ante-defined norms, which are precise ideas about what a specific test score means in relation to a content-psychological defined specific criterion. This method is called *criterion-oriented interpretation*.

As in the frame of the present research it is not yet intended to make statements about characteristic values of individuals or to interpret them with respect to a reference group, a calibration is not performed at this point. Thus, no rating scales to interpret the test scores were formed and no standard of comparison was defined.

As illustrated in the preceding sub-sections, manifold methodological decisions had to be made to develop the DAC-test. The sub-sections are supposed to give an overview of the procedure and clarify corresponding psychological vocabulary. The results of reliability and validity calculation are shown in section 5.2.

In addition to the six phases of psychometric test development, the next section contrasts the considerations for choosing the strategy of data collection with criteria for choosing the target group in sub-section 3.3.1, and the actual parameters of data gathering in sub-sections 3.3.2 and 3.3.3.

3.3 DATA COLLECTION STRATEGY

The data collection strategy was formed by two main objectives: achieving good quality data and receiving a great number of participants. Thus, the consideration around the data collection focused on how to get access to motivated test participants by thinking about tailoring the test design to the preferences of this group, communicating effectively with potential examinees, and providing an easily accessible and administered testing instrument – and consequently, reducing perceived barriers for test participants.

In terms of the test design a single unrepeated online survey for individuals was chosen to gather performance data on decision-making abilities (cf. section 4.1), data on self-assessments of the validation criteria (cf. section 5.1) and, data on socio-demographic variables (cf. appendix section 9.10). The mass survey was designed for self-completion. The main reason for choosing online testing was its advantage of gathering data from a large number of people within a short period of time, independent of their geographical location with relatively low resources (also cf. sub-section 3.2.1). Another reason was that online testing results are available very quickly for subsequent analyses (e.g. Dillman, Smyth, & Melani Christian, 2014). Even though the test content can be seen as complex with its eight dimensions of DAC, and the test length with its more than 60 minutes processing time as long, which does not create the optimal condition for Internet surveys (e.g. Robson, 2011), due to the limited resources, online testing appeared as the most promising option for conducting the present research.

In addition to the choice of online testing, a couple of other selection criteria for the target group appeared: Participating in the study was only possible if a person: (1) had access to the Internet and hence was able to receive the call for and the information about the study, and do the online test, (2) was familiar with operating a computer and, (3) was fluent in English.

In terms of approaching and communicating with the target group effectively, personalised e-mail (cf. Barron & Yechiam, 2002) requests, which included all relevant information (Dillman et al., 2014), were sent out from an e-mail that was exclusively established for the acquisition and communication with test participants (cf. appendix sections 9.4, 9.5, 9.6, and 9.7). Interested subjects received a reply within one working day with further details on the test procedure.

To better draw prospective examinees' interest in participating in the study and potentially raise their external motivation, several incentives were applied: First, direct feedback on an individual's test performance was provided right after finishing the test (e.g. Fitts & Posner, 1967) in the form of the achieved percentage test score for each DAC-dimension and the overall score. Second, participants were able to download a personalised certificate documenting their participation and their achieved test score. An example of the certificate with the test feedback is provided in appendix sub-section 9.9. Third, a monetary incentive (Bosnjak & Tuten, 2003) of €15.00 for participants of the main testing was offered. In general, it is assumed that the offered cash incentive effectively motivates participants to complete the test (Singer & Ye, 2013).

Transparency and discretion were a cornerstone of this study. So, before beginning the test, the following information was presented to all participants:

- The present DAC-test is part of the Ph.D. research of Nadine Oeser at the London School of Economics and Political Science.
- Participation is voluntary.
- The aim of the research project is to provide a valid and reliable psychometric performance test to measure Individual Decision-Making Competence.
- Participants will benefit in terms of individual feedback and a certificate with the test score.
- The DAC-test involves a performance test with several components of decision-making, self-rating and socio-demographic items.
- Refusal to participate may be given without reason. Participants may withdraw at any time without giving reasons[46].
- According to LSE's data protection rules, an appropriate analysis and saving of the test data are affirmed. The test data will only be used for research purposes.
- It is confirmed that every participant's anonymity is protected. Personal data will be saved separately from test scores. Personal information will be treated as strictly confidential and will not be made publicly available or given to a third party.
- Information generated by the study may be published. No details, from which participants could be identified, will be divulged, as only aggregated data (calculated over all test participants) will be used.

Additionally, participants were asked to agree to the use of their data for scientific purposes by consenting to the following sentence:

"By checking this box, I agree that my test results may be used for research purposes in the course of a PhD research at the London School of Economics and Political Science, UK".

The data collection strategies for pre-testing and main testing differed in terms of which channels were used to get access to potential examinees. The pre-testing (cf. sub-section 3.3.2) served, besides the item analysis, the purpose of finding the appropriate way to gain access to data. The main testing (cf. sub-section 3.3.3) benefitted from the insights from the pre-testing and was supported by financial resources for data gathering,

[46] Test participants were able to withdraw at any time as the test data was not saved until the DAC-test was completed. After completing the DAC-test, test participants received an automatically generated unique 15-digits code, saved independently from their test results. Together with an email, which was provided at the end of the test, test participants were able to withdraw from the study later on.

which were proportionately provided by the London School of Economics and Political Science and proportionately by the researcher.

Before sub-sections 3.3.2 and 3.3.3 illustrate the concrete procedures of data gathering in the pre- and main testing, the following section describes the arguments for choosing the target group.

3.3.1 Target Group

In order to define the *target group* for the present research, two selection criteria seemed to be relevant: Choosing a target group that... (1) is of social interest and therefore could help raise public attention to this research topic, and (2) goes along with a realistic chance, time- and moneywise, to gain access to the data of this group.

Corresponding to the two selection criteria, the group of adolescents and young adults was chosen as:

(1) Especially at this age the ability of making sound decisions can be seen as one of the key success factors of growing up. “Choices made in adolescence may have life-long consequences for the individual’s health, career, psychological well-being, and social acceptance.” (Mann et al., 1989, p. 265) Some studies characterise adolescents as poor decision-makers (e.g. Jacobs & Klaczynski, 2005): Eaton et al. (2010) have identified six risk factors in adolescence, including behaviours that contribute to unintentional injury and violence, alcohol and other drug use, or physical inactivity. Additionally, Slovic (2001) has shown that 80 out of 100 young smokers regret their decision to start smoking and would decide differently if they had the chance again. Also, career choices seem to lack careful consideration regarding long-term consequences leading to 6.0% of Germany’s teenagers leaving school before completing 9th grade (Berkemeyer et al., 2014). The rate of breaking off an apprenticeship is approximately 24.4% (Uhly, 2015). And, a quarter of Germany’s students leave universities or technical colleges without graduating (Heublein, Schmelzer, & Sommer, 2008). Targeting this group could raise the public attention for better educating young people in decision-making and preparing them for life (e.g. Baron & Brown, 1991; Jacobson et al., 2012; Weller, Moholy, Bossard, & Levin, 2014).

(2) There is a reasonable assumption that adolescents and young adults have a natural exposure to computers and the Internet (e.g. ARD Forschungsdienst, 2014). In fact, more and more people have broadband Internet access and “become [in general] more accustomed to completing various daily activities online” (Dillman et al., 2014, p. 301). This circumstance affects online testing positively (e.g. Couper, 2008) as people are increasingly completing online surveys. Additionally, studies show that on average 84% of people between 14 and 29 years own a computer or laptop and 95% access the Internet

at least once a day (Bertsch, Huth, & Arenz, 2011). Since the DAC-test is designed as an online test, it seemed worthwhile to choose a target group that is characterised by an experienced, almost natural handling of computers and webpages. As the majority of students age-wise belong to the target group and they are assumed to be a relatively easily-accessible group due to their flexibility in terms of time and their interest in research, it seems realistic to access data from this target group.

Even though a target group for this research starting from early adolescence would be of great interest, the LSE Research Ethics Committee adjudges potential research participants under the age of 18 to be possibly vulnerable. Thus, this research does not include study participants who did not confirm that they were 18 years old or older at the time of participating in the study.

According to definitions in developmental psychology (e.g. Zimbardo & Gerrig, 1996) older adolescents represent the age range from 18 to 21 years and young adults are approximately 20 to 30 years old. According to Jacobs and Klaczynski, (2005), Furby and Beyth-Marom (1992) and Reyna and Farley (2006), there are no concerns with significant age-related developmental differences in the decision-making of the two target groups, teenagers and young adults. “By mid adolescence, most individuals have approximately adults’ ... cognitive skills” (Fischhoff, 2008, p. 15). Consequently, people between 18 and 30 years were invited to take part in the study. In further chapters the target group is entitled with *young adults*.

3.3.2 Study Procedure of the Pre-testing

To enable the item analysis approximately 200 data sets were aspired (cf. Lienert & Raatz, 1998. p. 60). Thus, in June, July, and August 2014, in order to find potential test participants a group of the London School of Economics and Political Science summer school students taking the “Judgement and Decision Making for Management” course and voluntary students of the psychological department at Free University of Berlin and the management department of the University of Bayreuth were contacted via e-mail and invited to participate within an overall time-frame of two months (cf. e-mails in appendix sections 9.3 and 9.5).

People, who replied as willing to participate received an answer within one working day containing all necessary information about the duration of the study, privacy protection and the advantages of participation in advance.

After the actual pre-testing, three further ways to access data were tested: distributing flyers in Berlin cafés and bars, posting at social media platforms, for instance

Facebook, and acquiring test participants via online labour markets such as *Amazon Mechanical Turk*. Unfortunately, none of these three alternative acquisition strategies complied with both main criteria for the data collection strategy - obtaining a good quality of data and receiving a great number of participants. That is why the general procedure of data gathering was retained.

The following sub-section presents the strategy for data gathering for the main testing. The socio-demographics and results of the pre-testing are presented in section 4.2.

3.3.3 Study Procedure of the Main Testing

From June to August 2015, study participants for the main testing were obtained via the London School of Economics and Political Science summer school course "Judgement and Decision Making for Management", the course on "Management Systems Theory, Applications, and Design" from Virginia Tech, the Decision Analysis Society Newsletter, the Multi Criteria Decision Making mailing list, and the mailing lists of the Max Planck Institute for Human Development in Berlin, the Free University Berlin, and the University of Bayreuth. All potential participants were informed about the study details via e-mail, as was the case in the pre-testing (cf. e-mails in attachments 9.6 and 9.7).

As in the pre-testing, all examinees were informed in advance about the duration of the study, privacy protection and the advantages of participating in the study, including the receipt of their individual test feedback and the personalised participation certificate.

At the end of the planned period of one month for the main testing, only 34% of the intended 350 test participants had been acquired. Thus, a remuneration of €15.00 was offered as a financial incentive to acquire further test participants. Since the introduction of this external motivator may have an influence on the test results of the corresponding group of participants a variable to control for potential effects was included (for results cf. section 6.5.1).

"When a mixed-mode strategy is not possible, and e-mail is the only contact option, an electronic incentive sent to all sample members with the survey request is likely the best option." (Dillman et al., 2014, p. 331) Sending the remuneration via *PayPal*[47] was offered (Birnholtz, Horn, Finholt, & Bae, 2004). To counteract that people might not participate in the study because they did not have a *PayPal* account or would not want or know how to create one, transferring money to their bank account was offered as well.

[47] Paypal.com is an online money transfer service.

Since the risk that people would be reluctant to reveal their account information needed to be considered, the PayPal option was the option of first choice.

After completing the DAC-test test participants received automatically generated unique 15-digits code, saved independently from the test results to ensure confidentiality. To receive the €15.00 remuneration, test participants had to send the code to the conductor of the test and with it their PayPal or bank account information. The money was transferred within one working day.

As this chapter has given an overview of the standard procedure of psychometric test development according to e.g. Lienert and Raatz (1998) or Moosbrugger and Kelava (2012), explicated the methodological choices which were made, and presented the data collection strategy, the following chapter illustrates how the DAC-test was precisely constructed. It shows how each DAC-dimension was operationalized (cf. section 4.1) and what the process and the results of the pre-testing look like (cf. section 4.2).

The socio-demographics and results of the main testing are presented in section 5.2.

4 DEVELOPING THE DAC-TEST

Providing a psychometric performance test for capturing the cognitive dimensions of DAC (cf. Table 2-2) constitutes the main aim of this research. The present chapter builds the core of this Ph.D. research as its sections and sub-sections give a detailed compilation of the various segments of the constructed DAC-test. The chapter starts with considerations about the test's content and layout as well as subset information-technological frame conditions (cf. section 4.1). The chapter continues with sub-sections 4.1.1 - 4.1.8 demonstrating how each DAC-dimension was operationalized. Section 4.2 and its corresponding sub-sections present the process and results of item and distributional analyses of the pre-testing.

In the framework of the operationalization process of the eight cognitive DAC-dimensions, it had to be decided in which content-context the DAC-test should be embedded. At this point it was important to take one essential test constructional consideration into account: To support the analysis of statistical properties, the plan was to assess the reliability of the DAC-test by *parallel-forms reliability* (cf. sub-section 3.2.4). Therefore, two cases of equal structure have to be tested and analysed in the phase of *item analysis* and *distributional analysis* (cf. section 3.2.3). In comparison to split-half reliability[48], the advantage of parallel-forms reliability is that the two cases are considered as equivalent measures and therefore could be used independently. An additional practice-relevant benefit behind the idea of providing two cases was to enable pre-post-testing for later applications of the DAC-test.

Consequently, as the DAC-test aims to present a set of content-wise connected items, two complex real world decision cases, which require the performance of all cognitive DAC abilities, had to be designed. In discussions with experts from decision analysis, the topics of *career choices* and *investments* were identified as areas with major decision opportunities, which most young people and so the target group (cf. sub-section 3.3.1) experience at least once in their adolescence and early adulthood. The topics were chosen as they both fit Matheson's and Howard's (1968) characteristics of a complex decision problem (cf. sub-section 3.2.1) and cover a concrete decision, which the target group had to handle not too far in the past, is facing right now, or will deal with very soon. Thus, building upon the decision characteristics of Matheson and Howard, two

[48] Split-half reliability is applicable if a test contains a great number of comparable items, which can randomly be split into two comparable test parts.

decision cases broadly targeting the questions of *which vehicle to buy* (henceforth: INVEST case) and *what to do after having completed school* (henceforth: EDU case) were developed. In the EDU case, it is assumed that all test participants already had experienced this particular decision at least once. Consequently, the condition of all test participants in terms of this decision experience is presumed to be comparable. In the INVEST case, it is presumed that test participants either have experienced such a situation already by themselves or knew someone, e.g. partner, friends, family member, who were confronted with such a decision. Therefore, the INVEST case topic is supposed to build a well-accessible and realistic basis for the test, with which all test participants can identify.

Thus, the first version of the DAC-test presented two parallel cases. The INVEST case was set first, followed by the EDU case.

Preceding the operationalization of DAC in sub-sections 4.1.1 - 4.1.8, the next section argues the thoughts on test and item layout, including several information-technological considerations, and defines the eight cognitive DAC-dimensions.

4.1 OPERATIONALIZING DECISION-ANALYTICAL COMPETENCE

The construction of the psychometric performance test demanded two designing foci, which are summarised in this section and the following sub-sections: On the one hand, it had to be decided how to design the test and task layout taking into account advice about web page and test design. On the other hand, it had to be determined how to operationalize each dimension of DAC.

Accessibility and display

In order to provide the DAC-test to a broad population online enabling participants to complete the survey on a computer or a mobile device, the test was configured as a browser-based version optimised for mobile users. A consistent page layout was chosen to support test participants to easily process the given information and focus on answering the items (Dillman et al., 2014). Figure 4-1 shows an example of the screen design of the DAC-test. Instruction, item stem and corresponding scales are highlighted in blue. This information is visually dominant on the page and allows test participants to centre their attention upon it.

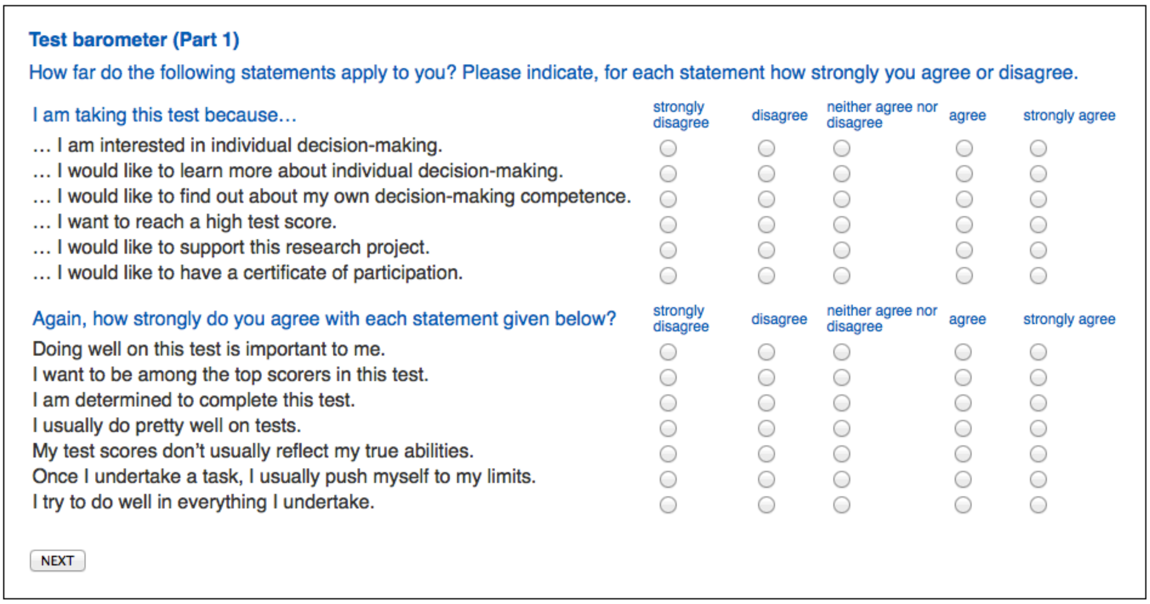
Items that appear very complex or whose display appears very long (for details cf. sub-sections 4.1.1 - 4.1.8) were arranged in a single-question-per-page design to enable test participants to concentrate on those more voluminous items and support focusing on every particular instruction. Even though tests with this design appear longer in terms of processing time to participants (e.g. Manfreda, Batagelj, & Vehovar, 2006), reducing the perceived complexity of the test (e.g. Dillman et al., 2014) appeared more important in the present case. To avoid test participants experiencing difficulties in remembering item relevant context given on a previous page, all necessary information was repeated on the actual page – either in the item stem or by a to-click-on example or definition.

According to Nielson (2004), checkboxes are used for lists of options where any number of presented choices could be selected. In contrast, radio-buttons are applied "when [there are] two or more options that are mutually exclusive and the user must select exactly one choice" (p. 1).

Test structure

After entering the welcome page, web page visitors have access to information on the test's and researcher's backgrounds. Additionally, they are informed about transparency and discretion issues (cf. section 3.3) and are given an overview of the various test parts and permitted testing aids. By clicking *start*, test participants are forwarded directly to the agreement of participation where they are asked if they are willing to continue with the test. In addition, test participants have to confirm that they are of legal age, i.e. 18 years or older, to ensure that they are legally permitted to give consent to their study participation. Additionally, they are asked to answer a few socio-demographic items, such as age, level of English skills, previous knowledge on decision science, and assess statements on their motivation and test intention. Before they actually start with the test, the following declaration appears:

"All of the mentioned situations and persons are entirely fictitious. Any similarities with existing ones are coincidental and unintended. Any mentioning of brand names or features typical of a brand should neither give the impression of endorsement nor refusal. Also, the tests, test results and statistics which are in relation to brand names have been completely invented."


Test barometer (Part 1)

How far do the following statements apply to you? Please indicate, for each statement how strongly you agree or disagree.

I am taking this test because…	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
… I am interested in individual decision-making.					
… I would like to learn more about individual decision-making.					
… I would like to find out about my own decision-making competence.					
… I want to reach a high test score.					
… I would like to support this research project.					
… I would like to have a certificate of participation.					

Again, how strongly do you agree with each statement given below?	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
Doing well on this test is important to me.					
I want to be among the top scorers in this test.					
I am determined to complete this test.					
I usually do pretty well on tests.					
My test scores don't usually reflect my true abilities.					
Once I undertake a task, I usually push myself to my limits.					
I try to do well in everything I undertake.					

NEXT

Figure 4-1 Showcase–page of the DAC-test layout (pre-test-barometer)

Test construction considerations

Before operationalizing the DAC-dimensions, two requirements for the to-be-developed indicators/items were defined to support a good testing experience for test takers, to encourage good quality of data, and to ensure efficient data analyses:

(1) Theory-driven and diversified: For the process of item creation a top-down approach was applied, meaning that specific attributes for operationalizing each DAC-dimension were collected from the decision-analytical literature. By constructing the items it was kept in mind that "[a] source of error in testing arises from boredom – especially in tests of ability and similar spheres, where effort and concentration are required. A variety of items is likely to make the test less monotonous for the subjects." (Kline, 2015, p. 47) Therefore, it was attempted to composite items of the various DAC-dimensions differently in terms of question type, scale type, given examples, and supporting graphics.

(2) Objective and efficient: As mentioned in section 1.2, the DAC-test aims to measure the quality of individual decision-making performance. Thus, each item was constructed in a way so that a specific target value was defined, which allows a direct assessment of the accuracy of each answer. In this context, the online test was programmed to automatically quantitatively analyse the answers and give each test participant feedback on the percentage achievement of the test score.

Control for data quality

To avoid item skipping and answer guessing, which would result in loss of data, test participants are systematically forced to answer each single item. If test participants try to proceed with the test after omitting one or several items on a page, a reminder in red letters appears at the top of the page that is linked directly to the unanswered item(s) requesting the participant to answer it/them in order to proceed. To reduce the chance of guessing correctly to true-false-items and other items with a small set of answer options, an *I do not know* answer option[49] was added (e.g. Muijtjens et al., 1999) with the explanation/hint: *Please omit an item by clicking "I do not know" only when you are convinced that your answer would be the same as a random guess*.

Information on how long it took participants to answer each item is collected and saved in the database automatically. Recording this data serves to control for the quality of data: Participants who click through the whole test and then wait for submission to pretend to take as long as thoughtful and serious test participants do, can be identified and excluded from analyses.

Before the following sub-sections introduce in detail how, for the pre-testing (cf. sub-section 3.3.1), each cognitive DAC-dimension was operationalized, Table 4-1 gives an overview of the eight DAC-dimensions on the basis of their definition (right column of Table 4-1). Each of the following sub-sections (left column of Table 4-1) delineates the operationalization of the corresponding DAC-dimension. Section 4.2 subsequently presents the results of the statistical analyses and correspondingly, the resulting changes to the operationalization of the dimensions for the final version of the DAC-test.

Table 4-1 Definitions of the eight DAC-dimensions

Sub-section	DAC-dimension	Abbr.	Definition
4.1.1	Ability to envision one's objectives	OBJECTIVES	This DAC-dimension describes the awareness of subjects of their individual preferences and the ability to articulate a clear set of personally relevant objectives during contemplation of an important decision (e.g. Hammond et al., 1999; Hastie, 2001; Keeney, 1992; Mellers & Locke, 2007).

[49] Formula scoring was applied to treat the *I do not know* answer option (cf. sub-section 3.2.2).

4.1.2	Ability to recognise decision opportunities	OPPORTUNITIES	This DAC-dimension describes the ability of subjects to identify decision situations, which are characterised by the potential to support decision-makers in achieving their desired goals, and to assess which degree of analytical thinking is needed to solve those situations (e.g. Keelin et al., 2009; Keeney, 1992).
4.1.3	Ability to assess decision fitness	FITNESS	This DAC-dimension describes the ability of subjects to decide whether emotions and physical status allow focusing on a concrete decision situation, proceeding analytically with the decision-making process, and making a conscious choice (e.g. Keelin et al., 2009; Thaler & Sunstein, 2008).
4.1.4	Ability to frame a decision	FRAME	This DAC-dimension describes the ability of subjects to perceive the relevant aspects of a decision situation and to interpret their meaning for the corresponding process of decision-making (e.g. Hastie & Dawes, 2010; Howard, 2007; Mellers & Locke, 2007).
4.1.5	Ability to identify relevant alternatives	ALTERNATIVES	This DAC-dimension describes the ability of subjects to envision alternatives that fit a set of given objectives in a decision situation (e.g. Goodwin & Wright, 2014; Hammond et al., 1999; Keelin et al., 2009; Keller & Ho, 1988).
4.1.6	Ability to deal with uncertainty	UNCERTAINTY	This DAC-dimension describes the ability of subjects to comprehend, interpret, and calculate probabilities of different qualities in order to make a choice (e.g. W. Edwards, 1954; Hastie & Dawes, 2010; J. E. Matheson & Howard, 1968; Raiffa, 1968).
4.1.7	Ability to integrate information	INFORMATION	This DAC-dimension describes the ability of subjects to combine assessments of uncertainties and values coherently, compare various alternatives and objectives, and correspondingly rank those alternatives in relation to their relevance for the decision-maker (e.g. Bazerman & Moore, 2009; W. Edwards, 1954; Goodwin & Wright, 2014; Keelin et al., 2009; Raiffa, 1968).
4.1.8	Ability to plan to implement a decision	IMPLEMENTATION	This DAC-dimension describes the ability of subjects to set up a strategy for how to translate a cognitive decision into action by formulating concrete next steps and thinking ahead about possible obstacles and ways to address them (e.g. Hammond et al., 1999; Keelin et al., 2009; Mellers & Locke, 2007).

Note. DAC = Decision-analytical competence; Abbr. = Abbreviation.

The order of the following sub-sections is geared to the order of appearance of the DAC-dimensions in the test.

4.1.1 Items of the Ability to Envision One's Objectives

The *ability to envision one's objectives* (henceforth: OBJECTIVES) is the first dimension of DAC presented in the test. This first part operationalizes the subject's awareness of their individual preferences and their ability to articulate a clear set of personally relevant objectives during contemplation of an important decision (e.g. Hammond et al., 1999; Hastie, 2001; Keeney, 1992; Mellers & Locke, 2007; Raiffa, 1968).

There were two reasons for positioning this task as the first dimension of the test:

First, as presented by Keeney's (1996) *value focused thinking*, envisioning one's objectives is a good starting point for making a decision as it supports decision-makers in realising what they actually want to achieve with their decision. Especially for complex decisions, the addressed decision type in this research, the corresponding task is assumed to be a suitable origin for the test as "Keeney's approach is particularly worth considering for major strategic or life-changing decisions where there is a need to think deeply about what you want to achieve in life" (Goodwin & Wright, 2014, p. 55).

Second, the task requires test participants to list as many personal objectives as possible for a desired concrete complex decision situation. Since the test parts following this item provide stories and examples around either the INVEST case or the EDU case, it was aimed to avoid the contents of the other DAC-test tasks influencing the test participants' answers for this particular item.

Test concept and theoretical basis

For the operationalization of the DAC-dimension OBJECTIVES it was possible to choose a *criterion oriented/external test concept* as the work of Bond, Carlson, and Keeney (2008, 2010) provides a well-developed approach to measure this ability. In their studies, the authors ask test takers to articulate a complete and clear set of relevant personal objectives for a concrete decision e.g. "What are your objectives for choosing an internship?" Bond et al. operationalize OBJECTIVES in four steps (cf. Figure 4-2): (1) Test takers have to generate as many relevant objectives as possible in a list and then put it aside. (2) Test takers receive a so-called *master list* with an exhaustive set of objectives. They are asked to check all objectives that are relevant to them. (3) Test takers take their list with self-generated objectives from step 1 and the master list from step 2 and are requested to match the self-generated objectives to the ones of the master list. (4) Test takers are asked to rank the importance of their (in step 2) checked objectives.

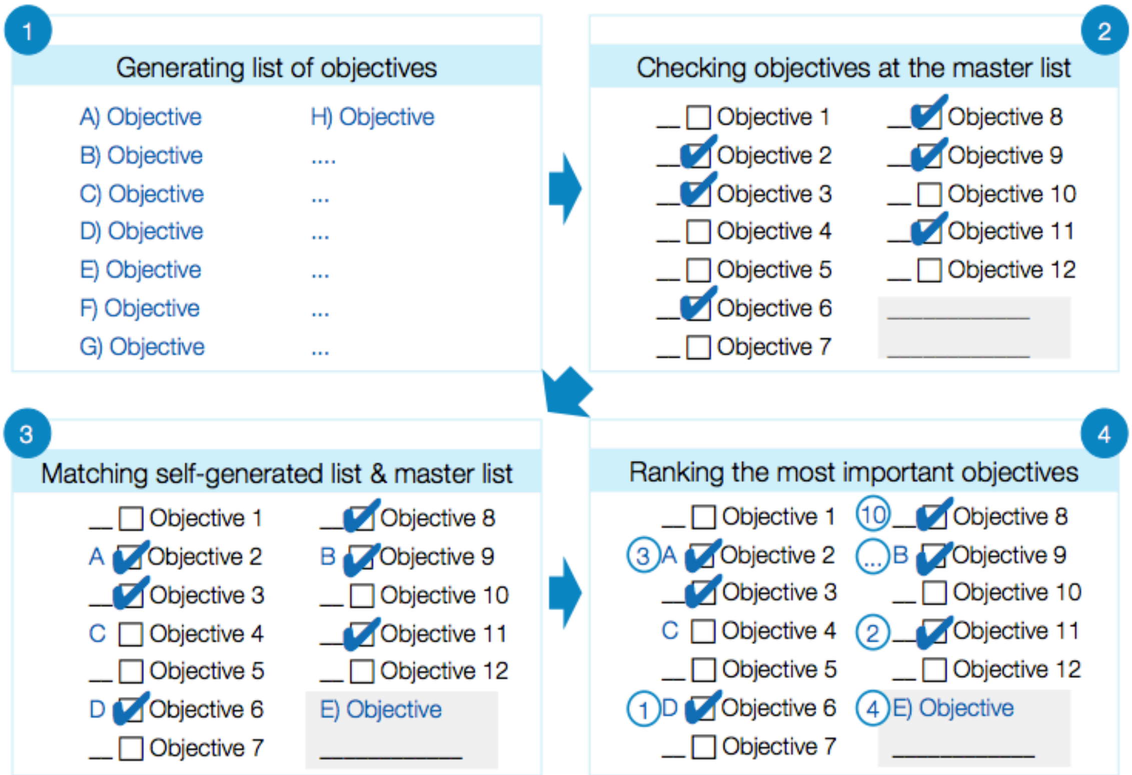


Figure 4-2 Outline of task procedure (cf. Bond et al., 2008, p. 59)

Operationalization and layout

For the present research, content-wise adapted items in comparison to Bond et al.'s item set are presented (cf. appendix sub-section 9.8.1). A single-item-per-page layout was chosen as the display of each step takes plenty of space. Additionally, it is important for measurement purposes that test participants solve this task step by step. For this reason, test participants are not able to back up. So their set of self-generated objectives from step 1 cannot be modified or complemented in retrospect.

The test starts with the instruction and the corresponding example item (cf. Figure 4-3) of the first step to operationalize OBJECTIVES:

Step 1: Imagine you have to decide what to do after finishing school. What would be your most relevant objectives for choosing a direction? Please list as many objectives as you can think of, writing each one in the lines (from A to AD) below (cf. example).

EXAMPLE ITEM

Imagine you are hungry and want to go out for dinner. Please list as many objectives as you can think of, writing each in one of the lines (from A to AD) below.

My objectives for choosing a restaurant are:

A) Get good value for money
B) Have vegetarian options
C) Be close to my current location
D) Offer nice atmosphere
E) Offer food, which is not too spicy
F) Offer friendly service
G) Offer fast service
H) ...

Figure 4-3 Operationalization of OBJECTIVES – example step 1

For the test participants it is possible to list in 28 lines up to 28 of their personal objectives. After having completed this task and clicking *next,* the instruction of the second step with its example item (cf. Figure 4-4) and a list of 43 potential objectives is presented as follows:

Step 2: Again, imagine after you finished school you have to decide what to do now with your life. Please select all objectives that appear relevant to you for selecting a direction for your life by ticking the checkbox on the left of each objective in the list below (cf. example).

EXAMPLE ITEM

You are sure that you are going out for dinner tonight. Please select all objectives that appear relevant to you for choosing a restaurant by ticking the checkbox on the left of each objective in the list below.

⦿ Offering good-quality food
⦿ Having reasonable prices
⦿ Being located in the neighbourhood
⦿ Offering a good atmosphere
⦿ Offering a good service
O Providing free parking
⦿ Having the chance to get a table for tonight without reservation
O Being accessible for handicapped
O Offering a separate smoker area

⦿: selected objective O: unselected objective

Figure 4-4 Operationalization of OBJECTIVES – example step 2

Test participants can select up to 43 objectives that appear relevant to them from the list with the presented potential objectives. The list of 43 objectives was compiled in a small pilot study (cf. appendix section 9.1). The main reason for creating such a master list was to provide the DAC-test participants a list that contains all relevant possible objectives. An overview of the list of potential objectives can be found in appendix sub-section 9.8.1.

On the next page both lists – the list of the self-generated objectives from step 1 and the list with the 43 potential objectives from step 2 – are displayed with the following instruction and example item (cf. Figure 4-5):

Step 3: Please match each objective you listed in the first task (now displayed on the left side below) to the objectives on the right side by writing its letter to the left of them. If some of your personal objectives do NOT match any objective here, please write them down in the shaded area below.

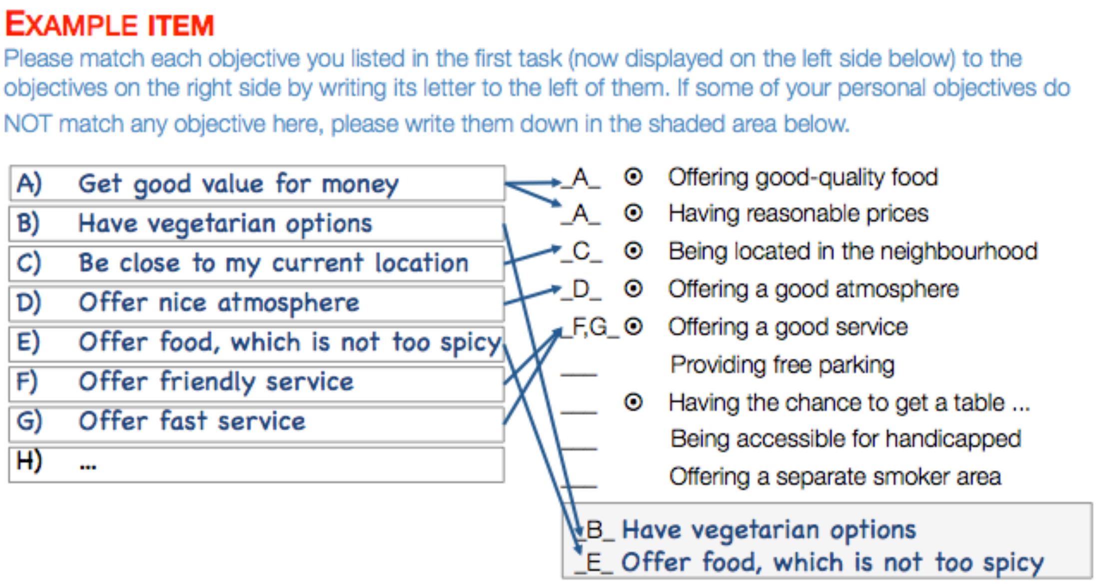


Figure 4-5 Operationalization of OBJECTIVES – example step 3

To proceed to the final step of this task, test participants have to match their personal objectives with the given list of objectives and add self-generated objectives that could not be matched. The matching process is necessary to gain clarity and structure of both self-generated and listed objectives for participants and the test analysts. Therefore, a constructive basis for the final step of the task is built and a clear assignment of personal formulations and presented descriptions of objectives received.

The fourth and final step of this task consists again of a selecting task. Test participants have to select their seven[50] most important objectives for the presented decision situation containing self-generated and/or recognised objectives[51]. The corresponding instruction with its example item (cf. Figure 4-6) is illustrated in the following:

Step 4: Finally look at all objectives you have selected in the second task and the ones in the shaded area. Please mark the 7 most relevant ones for you by ticking the checkbox on their very left side.

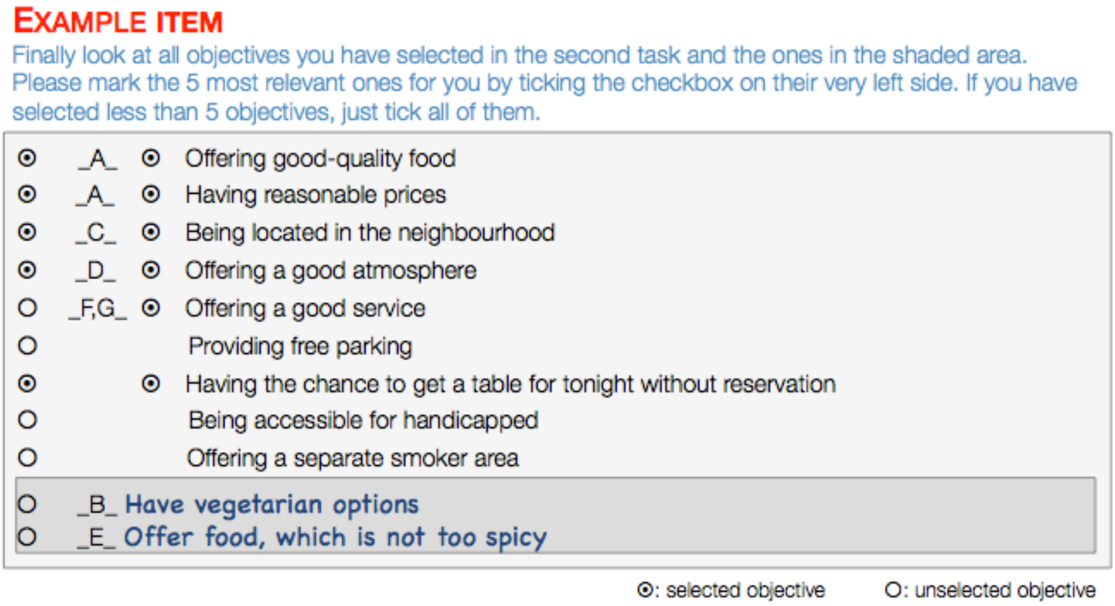


Figure 4-6 Operationalization of OBJECTIVES – example step 4

Scoring

"In decision theory, values are a matter of individual taste. Their accuracy cannot be evaluated in terms of an external standard." (Parker & Fischhoff, 2005, p. 4) That is why a specific evaluation criterion had to be defined for this task of the DAC-test. The higher the percentage of self-generated objectives of the final selected seven objectives, the better the OBJECTIVES. For the scoring the number of self-generated objectives of the selected

[50] According to decision scientists and practitioners such as Ralph L. Keeney decision-makers face on average a set of six to 10 (R. L. Keeney, personal communication, June 28, 2016) fundamental objectives when solving a complex decision problem. As in the framework of the DAC-test, test participants might feel time pressure when solving the tasks, the magical number seven of Miller (1994), which refers to the number of objects an individual is able to retain in working-memory, was set as a threshold for a complete set of relevant objectives.

[51] While the term "self-generated objectives" describes objectives that have been listed by the test participants in step 1 of this block of tasks, "recognised objectives" define objectives chosen in step 2 of this block of tasks as relevant by the test participants from the master list of objectives.

seven most important ones was counted. Thus, for the pre-testing a maximum score of seven points could be reached for OBJECTIVES.

4.1.2 Items of the Ability to Recognise Decision Opportunities

The second cognitive DAC-dimension being measured is the *ability to realise decision opportunities* (henceforth: OPPORTUNITIES). It describes the ability of subjects to identify decision situations, which are characterised by the potential to support decision-makers in achieving their desired goals and in assessing which degree of analytical thinking is needed to solve those situations (e.g. Keelin et al., 2009; Keeney, 1992).

Test concept and theoretical basis

For the DAC-dimension OPPORTUNITIES the literature does not provide an existing approach to operationalize this skill. Important within this DAC-dimension is that a good decision-maker is able to differentiate between situations that provide a chance to decide and thereby consciously influence future events, and situations that do not offer such an opportunity (Keeney, 1996). In this sense, a decision situation describes a "choice of action – of what to do or not to do" (Baron, 2008, p. 6) that is driven by decision-makers' motivation to achieve their goals. A decision situation is characterised by "more than one possible course of action" (Hastie & Dawes, 2010, p. 24). As the DAC-test is ascribed to complex decision situations (J. E. Matheson & Howard, 1968), OPPORTUNITIES also captures the ability to differentiate between simple and complex choices, i.e. between decision situations, which require either a more rapid and automatic or a more deductive and controlled handling. The *dual-process* theories distinguish between two different cognitive processes or action strategies – intuition and reasoning (e.g. Chaiken & Trope, 1999; Sloman, 1996). Whereas the so-called *system I* is intuitive, impulsive and emotional, reacts fast to a stimulus, and is automatic, effortless and associative, the so-called *system II* is explicit, logical, controllable and slower but flexible in learning (Kahneman & Frederick, 2002). Consequently, for effectively dealing with more complex decision problems, the reasoning *system II* appears to be more suitable (Bazerman & Moore, 2009).

Operationalization and layout

Regarding theoretical decision-analytical definitions on how a decision situation is described, mini cases were constructed. The mini cases, each with four to seven sentences, vary in terms of presenting a decision versus a non-decision situation. Non-decision situations are characterised by, for example, missing alternatives (Howard, 2007). The cases that describe a decision situation differ in terms of the decision types presented, i.e. the character described in the case is confronted either with a complex or a simple decision or it is indistinct. While complex decisions are defined for example as costly and uncertain (J. E. Matheson & Howard, 1968, p. 3) requiring an analytical decision-making process, simple[52] decisions describe frequent/periodic choices for which decision makers have clear preferences, (cf. Keelin et al., 2009). Hence, test participants have to decide not only whether the character is facing a decision or not, but also which kind of decision type is presented and correspondingly the potential impact on the character's life. For the DAC pre-testing, eight items of the first type and four items of the second type were constructed. The list of mini-cases of the final DAC-test is presented in the appendix sub-section 9.8.2.

At the top of the page the instructions with the corresponding scale explanations (cf. Figure 4-7 and Figure 4-8) are presented:

Please read the following case descriptions carefully and decide whether the character is facing a concrete decision situation or not by ticking the appropriate button.

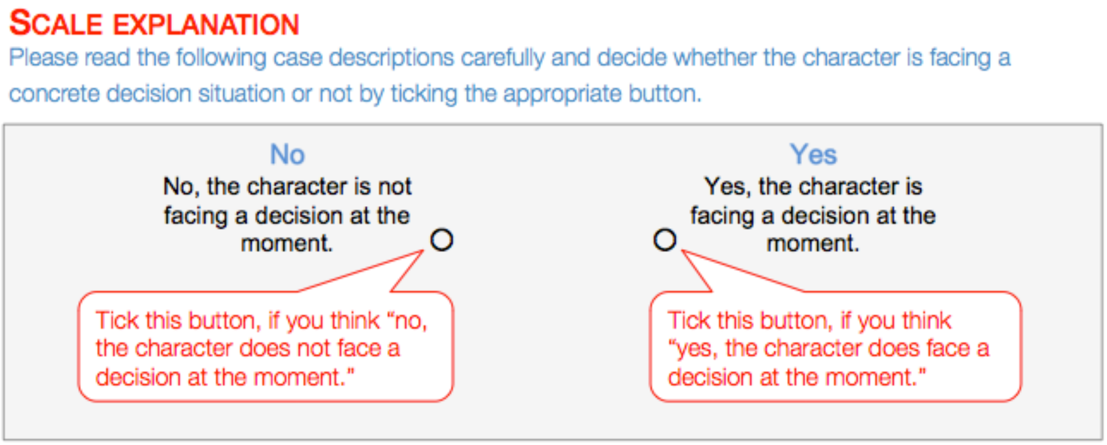


Figure 4-7 Operationalization of OPPORTUNITIES – scale explanation 1

If you think that the character is facing a decision situation, an additional question will appear. If this is the case please assess the potential impact of this decision on the character's life by ticking the appropriate button.

[52] To avoid test participants interpreting the word "simple", the wording "big vs. small decision" is used in the test.

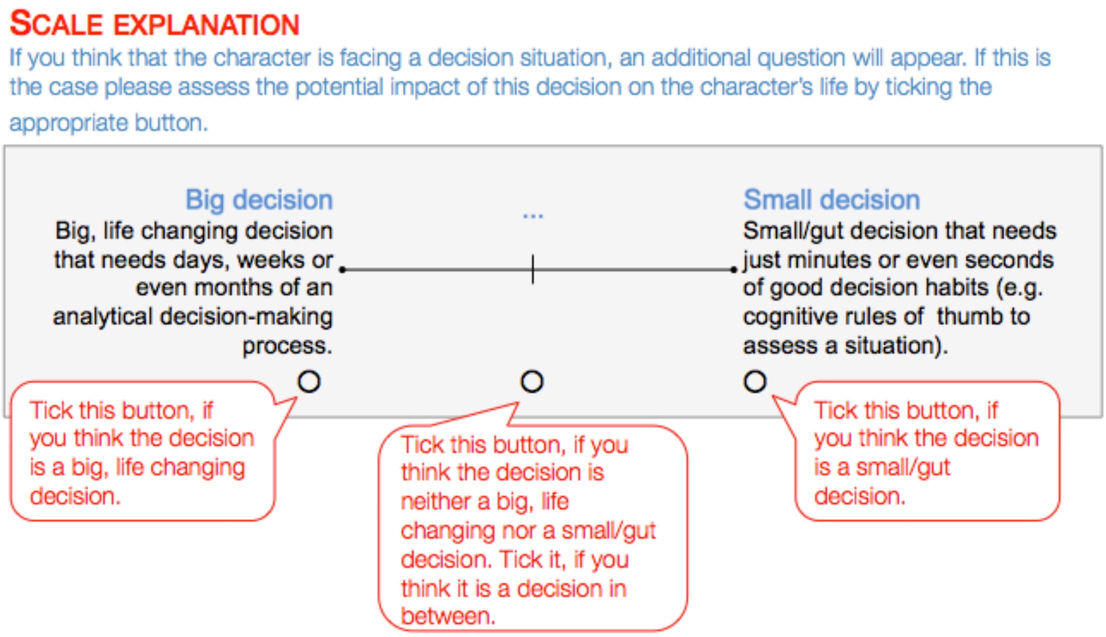


Figure 4-8 Operationalization of OPPORTUNITIES – scale explanation 2

The instructions are followed by the list of mini cases with the corresponding scale(s) all on one page. An example item for OPPORTUNITIES looks like the following:

If everything goes well, Luca will obtain his school-leaving qualifications next year. He has talked a lot with his parents about what he wants to become when he is grown up. Now he is thinking about what to do after passing his exams. Tomorrow, he will be able to apply for study programmes online, using his last school report.

○ No ○ Yes ○ I do not know

○ Big decision ○ ○ Small decision ○ I do not know

Scoring

On both scales an *I do not know* option is applied to reduce the chance of guessing correctly. In the case of the example item above, the correct answers would be “Yes” and “Big decision”. In terms of scoring, binary data is calculated, i.e. only the correct answer scores with one point per correct answer. The option *I do not know* and wrong answers do not score. So, for the pre-testing a maximum score of 12 points could be attained for OPPORTUNITIES.

4.1.3 Items of the Ability to Assess Decision Fitness

The third dimension of the DAC-test is the *ability to assess decision fitness* (henceforth: FITNESS), which describes the ability of subjects to decide whether emotions and physical status allow focusing on a concrete decision situation, proceeding analytically with the decision-making process, and making a conscious choice (e.g. Keelin et al., 2009; Thaler & Sunstein, 2008).

Test concept and theoretical basis

As for the previous dimension of DAC, prior research does not provide an approach to measuring FITNESS. On a theoretical level, according to Keelin et al. (2009), decision fitness describes the level of a person's sound decision-making habits for complex decision situations. Keelin and colleagues describe four steps to becoming decision fit: First, decision-makers have to learn to differentiate between physical and emotional statuses, which relate to reasoning when making a decision and inability to use analytical decision processes. Second, decision-makers have to "[g]ain a deep understanding of and skill for making quality decisions so as to rapidly go through" (p. 18) the steps of an analytical decision-making process. Third, decision-makers have to train and practice their decision skills in various situations. Fourth, the more often decision-makers apply the learned decision-analytical steps for sound decision-making, the sooner they develop good decision habits, which then merge in terms of the required cognitive system - from *system II* to *system I* (Kahneman & Frederick, 2002).

As the first step of Keelin et al.'s (2009) approach to becoming decision fit circumscribes the construct the DAC-test intends to measure, the first cognitive component of Keelin et al.'s (2009) approach is selected to be operationalized as FITNESS - knowing the difference between being decision fit and unfit.

Operationalization and layout

Subsequent to operationalizing FITNESS in the framework of the present research, an indicator had to be found that measures how far individuals recognise whether a decision-maker's physical condition and emotional status allow the decision-maker to run an analytical decision-making process. Therefore, situational judging tasks were developed. Situational judgment tasks/tests "are [usually] personnel selection instruments that present job applicants with work-related situations and possible responses to the situations." (McDaniel, Hartman, Whetzel, & Grubb, 2007, p. 63) For the present research, this kind of test instrument was slightly adapted, so that decision situations were presented to test participants. In this context, mini cases with five to 10 sentences describing career choice situations were constructed, for which test participants had to decide whether the character in each case is in the right mood and/or physical condition to analytically address the decision faced. Circumstances, such as being very angry, being under time pressure, having missed a proper amount of sleep, or being enraged or being extremely exhausted, shall indicate to the test participants that the character does not meet the requirements for a sound analytical decision process (Keelin et al., 2009).

As with the items of OPPORTUNITIES, the mini case items of FITNESS and the corresponding scale(s) are displayed in a list on one page. The following instruction with the corresponding scale explanation (cf. Figure 4-9) is displayed at the top of the web page:

Please read the following case descriptions carefully and pay attention to the character's physical condition and emotions. Please decide whether it is advisable to make a decision in his or her situation. For each case, please tick the appropriate button.

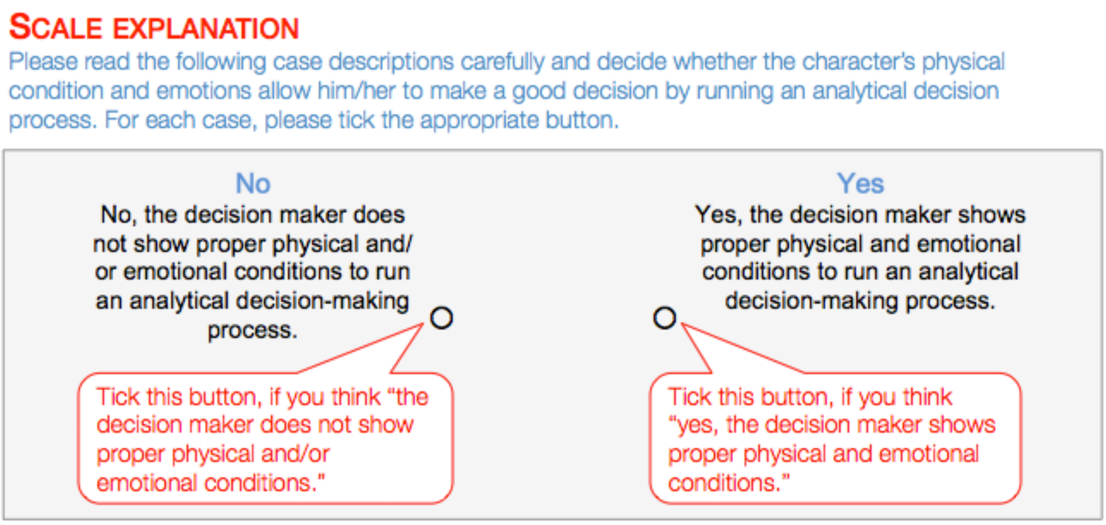


Figure 4-9 Operationalization of FITNESS – scale explanation

For the DAC pre-testing, eight items of this type were developed. The list of mini-cases of the final DAC-test for this dimension is presented in the appendix sub-section 9.8.3. The following example item serves to convey an impression of how FITNESS is operationalized:

Susan's parents expect her to become a dentist, like her father. Susan however would rather study literature. Tonight they had a heated argument on this topic. Her parents refuse to finance her "foolish ideas" and Susan shouted back at them. She ran back to her room, slammed the door and thought: "I will show them how I can study literature. I don't need their help! I will complete my online application now."

○ No ○ Yes ○ I do not know

Scoring

Comparable to the items of the preceding DAC-dimension, an *I do not know* option is provided to decrease the probability from 50% to 33.33% for randomly choosing the correct answer. According to the scoring for OPPORTUNITIES, binary data is calculated in the same style with one point per item for each correct answer. The correct answer for the example item presented above would be "No". The option *I do not know* and wrong answers do not score. Thus, for the pre-testing a maximum score of eight points could be attained for FITNESS.

4.1.4 Items of the Ability to Frame a Decision

The fourth block of DAC-tasks operationalizes the *ability to frame a decision* (henceforth: FRAME). It is defined as the ability of subjects to perceive the relevant aspects of a decision situation and to interpret their meaning in the corresponding process of decision-making (e.g. Hastie & Dawes, 2010; Howard, 2007; Larrick, 2009; Mellers & Locke, 2007)

Test concept and theoretical basis

According to Larrick (2009), sound decision-making involves the consideration of a broad decision frame. This includes taking "into account (1) multiple objectives – not just the most salient one at the moment; (2) multiple alternatives – not just the first option that lands on the table; and (3) multiple outcomes that could arise in the near and long term – not just the expected state of the world." (p. 461) To successfully solve a decision problem, decision-makers are required to ensure that their understanding of the situation is complete. This builds a basis for accurate judgments (Soll & Klayman, 2004).

According to the decision-analytical literature, no measure to operationalize this ability is provided. However, Bruine de Bruin et al. (2007) and Parker and Fischhoff (2005) have addressed this ability from the behavioural decision science perspective and thus, by the ability to resist decision biases and not giving in to framing effects.

Operationalization and layout

As the present research intends to measure DAC and thereby relies on the decision-analytical avenue to decision-making, the items of Bruine de Bruin et al. (2007) and Parker and Fischhoff (2005) did not seem to be sufficiently capturing this approach. Thus, a main case, which builds in terms of content the basis for the other up-coming DAC-test tasks, was constructed. It presents a character facing a complex personal decision including all relevant aspects of the situation, such as various alternatives (e.g. Hammond et al., 1999; Mellers & Locke, 2007), multiple objectives (e.g. Goodwin & Wright, 2014; Keeney, 2007) and uncertainties (e.g. Goodwin & Wright, 2014; Hastie & Dawes, 2010). The main case consists of approx. 750 words written in prose containing direct and indirect speech.

Besides including decision-relevant information such as objectives and alternatives, the case also gives details on secondary – partially superfluous – facts, such as relatives of the decision-maker or the location of the decision. This kind of additional information is given for two reasons: First, the case aims to present a concrete decision situation, which all test participants could face at least once in their lifetime, as realistically as possible; so background information and small details are given. Second, the secondary information serves as distractors (cf. sub-section 3.1.2; Kline, 2015) for the actual task to reduce the chance of guessing correctly. After having read the case study, participants are asked to click *next* to continue. The challenge for test participants is to differentiate between objectives and conflicting objectives, to recognise present alternatives, and to identify uncertainties in this concrete decision situation.

The main case fills almost one screen of an average laptop screen of 13" or 15". The main case of the final DAC-test is presented in sub-section 9.8.4 of the appendix. Following the presentation of the main case, the task starts with the following instruction:

Please have a look at the following list of taken notes from the case and decide which kind of information it is in the decision-making situation by choosing the most appropriate description from the drop-down menu for each fact.

Presented on one page, test participants have to assign facts from the main case to the appropriate description in this decision-making situation by selecting the compatible descriptions from a provided drop-down menu. A list of facts/notes from the main case in the form of bullets is presented in a column on the left side of the screen. On the right side of the screen a drop-down menu for each fact is presented. The menu contains the descriptions of alternative(s), conflicting objectives, decision-maker, family, friend, location, objectives, resources, time frame, and uncertainty. Overall, 12 items of this item type were created for the DAC pre-testing. Out of the 12, four items served as distractors. The list of items and items that served as distractors of the final DAC-test is depicted in sub-section 9.8.4 of the appendix.

Scoring

To score, the answer of an item has to satisfy two criteria. First, only the correct mapping of the following four descriptions score: objectives, conflicting objectives, alternative(s), and uncertainty. The other descriptions serve as distractors. Second, only if test participants select the correct/requested description from the drop-down menu, their solution scores - for instance assigning the fact *saving money and travelling* to the description *conflicting objectives*[53]. The scores are saved as binary data – correct vs. not correct. So, for the DAC pre-testing a maximum score of eight points could be reached for FRAME.

4.1.5 Items of the Ability to Identify Relevant Alternatives

The fifth dimension being operationalized in the DAC-test is the *ability to identify relevant alternatives* (henceforth: ALTERNATIVES) representing the skill of subjects to envision alternatives, which score high on a set of given objectives in a decision situation (e.g. Goodwin & Wright, 2014; Hammond et al., 1999; Keelin et al., 2009; Keller & Ho, 1988).

Test concept and theoretical basis

The decision-analytical literature provides different methodological approaches to generating alternatives (e.g. Keeney, 1992; Keller & Ho, 1988) and empirical research on procedures to create alternatives (e.g. Jungermann et al., 1983; Pitz, Sachs, & Heerboth, 1980; Selart & Johansen, 2011). The latter type of studies descriptively examines which kind of procedural changes may influence the quantity and quality of generating alternatives. While alternatives' quantity is determined by the number of alternatives, alternatives' quality is evaluated in terms of creativeness. Studies such as Gettys, Pliske, Manning, and Casey (1987) show that the generation of a broad set of alternatives satisfies decision-makers, independent of how many relevant options they were able to come up with. As in effective decision-making processes it is necessary to understand how well the alternatives satisfy the corresponding objectives - as "[a] decision can be no better than [the] best alternative." (Hammond et al., 1999, p.7) - the qualitative aspect of identifying alternatives is the focus of the present research.

Operationalization and layout

The most recent research of Siebert and Keeney (2015) provides a first attempt to assess the quality of a person's ability to identify alternatives by test takers themselves. Their approach is geared to the procedure of Bond et al. (2008); cf. sub-section 4.1.1) of measuring how well decision-makers are able to envision their objectives. In Siebert's and Keeney's study, test takers listed as many alternatives as possible in a first step, selected relevant alternatives from a master list in a second step, matched their self-generated list with the master list in the third step, and finally evaluated the quality of their selected alternatives on criteria such as the suitability to achieve given objectives. Even though the method of Siebert and Keeney could have been adapted and used for the present

[53] For more background information on the case cf. appendix sub-section 9.8.4.

research, it did not seem to be the best way to operationalize this DAC-dimension because of one crucial consideration: In contrast to the quality of personal objectives, for which the setting of an external and objective evaluation standard is not possible (Parker & Fischhoff, 2005), it is entirely possible to set such an impartial evaluation standard for the quality of alternatives. So, decision facilitators endeavour to enable decision-makers to recognise and verbalise every relevant objective (e.g. Phillips & Phillips, 1993; Phillips, 2007), but they cannot be clear on whether the articulated set of objectives is complete. In contrast, an external person can judge whether an alternative matches corresponding objectives (cf. Keeney, 1992). For this reason, an objective, non-manipulable way of assessment, in comparison to an assessment by the test taker, was desired. Therefore, in the framework of the present research, new items to capture this ability have been developed.

Thus, it was intended to operationalize the ALTERNATIVES' qualitative aspect by assessing alternatives according to their fit to objectives. In other words, test takers are asked to assess whether alternatives score on the presented objectives. So by a matching task, they have to evaluate the so-called fit of alternatives to corresponding decision objectives.

The complete item pool of ALTERNATIVES with its instruction is laid out on one page. The task starts with three sentences introducing the story of a character who is facing a concrete decision situation. The three objectives of the character are presented and test participants are instructed (cf. corresponding scale explanation in Figure 4-10):

Knowing about the character's objectives, please choose which of the following alternatives fit his/her three objectives all at the same time. Please tick the appropriate buttons.

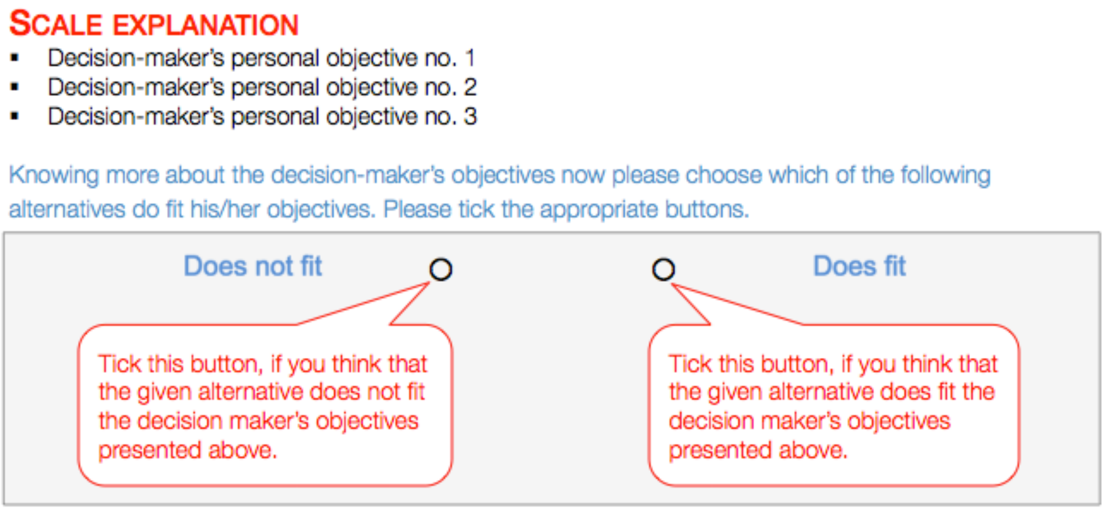


Figure 4-10 Operationalization of ALTERNATIVES – scale explanation

For the DAC pre-testing of this DAC-dimension an item set of 10 items was constructed. An overview of the item set of the final DAC-test is given in appendix sub-section 9.8.5. An example of an item for ALTERNATIVES looks like the following:

Here is what Brad thinks is relevant regarding his decision:

- I am really happy to be out of school now. I am not going to study straight away – Forget it!
- Whatever it is – I need to get some money for living. My parents aren't going to pay.
- I have always liked the countryside. Which alternatives are out there that let me spend some time outside?

Alternative	Scale		
Volunteering in an orphanage	○ Does not fit	○ Does fit	○ I do not know

Scoring

The scale was designed so that the presented alternatives scored either on one, two, all or none of the presented three objectives. The main aim of this decision task was to identify the alternatives that "fit" all objectives. Subsequently, those alternatives had to be categorised as "does fit"-alternatives. In case of alternatives scoring just on one two or no objective, test participants had to select "does not fit". For this DAC-dimension, an additional *I do not know* option was added to the scale (*does fit – does not fit)* to reduce the probability of subjects randomly choosing the correct answer. Binary data is calculated for the scoring. Thus, each correct answer scores with one point per item. The option *I do not know* and wrong answers did not score. The correct answer for the given example would be "Does not fit". Thus, for the DAC pre-testing a maximum score of 10 points could be reached for ALTERNATIVES.

4.1.6 Items of the Ability to Deal with Uncertainty

The *ability to deal with uncertainty* (henceforth: UNCERTAINTY) is the sixth dimension of the DAC-test. It describes the ability of subjects to comprehend, interpret, and calculate probabilities of different qualities in order to make a choice (e.g. W. Edwards, 1954; Hastie & Dawes, 2010; J. E. Matheson & Howard, 1968; Raiffa, 1968).

Test concept and theoretical basis

Uncertainty and similar constructs such as ambiguity or risk form major concepts in decision science literature (D. E. Bell et al., 1988; W. Edwards, 1954; Lipshitz & Strauss, 1997). This attention is justified as uncertainty builds a key obstacle to successful decision-making (e.g. Orasanu & Connolly, 1993). To overcome this obstacle, the attributes for good decision-makers are: thinking probabilistically and making judgment on systematically collected data (Hastie & Dawes, 2010; Kahneman & Lovallo, 1993). Decision science provides various approaches of how people cope with uncertainty; either by reducing uncertainty (e.g. Dawes, 1988; Hirst & Schweitzer, 1990), suppressing uncertainty (e.g. Boleman & Deal, 1991; Montgomery, 1988), or acknowledging uncertainty (e.g. Coombs, Dawes, & Tversky, 1970; Raiffa, 1968). The latter avenue of coping refers to the *rational choice theory*, in which a strategy to systematically include uncertainty into the evaluation of decision alternatives is applied. According to the *rational choice theory*, it is assumed that individuals have preferences among a set of decision alternatives. The attractiveness of an alternative is defined by the function of three components: the preference for its outcome, the probability that it will appear, and its costs (Raiffa, 1968). One of the most common techniques to identify the alternative that leads to the highest expected value is decision tree charts (Hastie & Dawes, 2010).

Operationalization and layout

Drawing from probability theoretical literature (e.g. Kolmogorov, 2013), it was intended to operationalize those aspects that are essential for making decisions under uncertainty. On the assumption that good decision-makers are able to comprehend and interpret probabilities, calculate expected values and conditional probabilities (cf. sub-section 2.4.2), textbooks and related decision scientific books (e.g. Gigerenzer, 2003; M.

Peterson, 2009; Shaughnessy, 2009) were reviewed in order to identify probability tasks that could be adapted to the topic of the INVEST case and the EDU case. In this context, the test was intended to cover the following areas of uncertainty tasks: perceive and comprehend probabilities, interpret their meanings, calculate basic probabilities as well as sensitivity and specificity, solve random experiments, and calculate expected values and natural frequencies (cf. Hoffrage & Gigerenzer, 1998).

For the pre-testing, eight different probability tasks were selected (cf. Table 4-2) - covering different levels of item difficulty according to the textbooks. In order to lead test participants to the more difficult items of probability calculation and to keep them motivated, the item block of this DAC-dimension was graded. The block of tasks starts with the easier items and closes with the more complex ones. Therefore, the complexity of all item types was analysed by determining the number of steps of thought, item by item: the fewer the steps of thought the less complex the item.

Table 4-2 Overview of UNCERTAINTY items of the DAC-test (pre-testing)

Object of measurement	Points to be reached
Understanding the meaning of probabilities by comparing two probabilistic pieces of information	1
Calculating relative frequencies by gathering the relevant information from a table or figure	1
Comprehending the literal meaning of probabilistic information	1
Calculating a two-stage random experiment by calculating one conditional probability	1
Calculating a random experiment with stochastically independent events	2
Calculating expected values of two potential alternatives	1
Calculating sensitivity/specificity with percentages	1
Calculating sensitivity/specificity with natural frequencies	1

In contrast to the other DAC-dimensions presented so far, whose items are of the same structure, the items of this DAC-dimension structurally differ: different instructions with or without corresponding figure and/or term definition, different scales with and without scale explanations. The eight items are presented on three pages. The items and scales of the final DAC-test are presented in sub-section 9.8.6 of the appendix.

Understanding the meaning of probabilities by comparing two probabilistic pieces of information

The first item of UNCERTAINTY was set as an icebreaker item (Kline, 2015) and thus relatively simple. It intends to measure whether test participants understand the meaning of probabilities by comparing two probabilistic pieces of information. Therefore, a four-sentence decision case referring to the main case of the test is presented, providing the probabilities of events A and B in percentages. After reading the case description test participants are asked which alternative the character should choose to maximise his/her chance for the desired outcome. To answer this item test participants can choose between two given probabilities. Additionally, an *I do not know* option is provided.

Calculating relative frequencies by gathering the relevant information from a table or figure

The second item assesses whether test participants are able to calculate relative frequencies by gathering the relevant information from a table or figure. To operationalize this sub-skill of UNCERTAINTY, a five-sentence case and a figure containing the necessary information/numbers are presented. Figure 4-11 gives an example of how the second item of this DAC-dimension asks test participants to calculate the chances of winning for the blue (A) and the red (B) wheel of fortune.

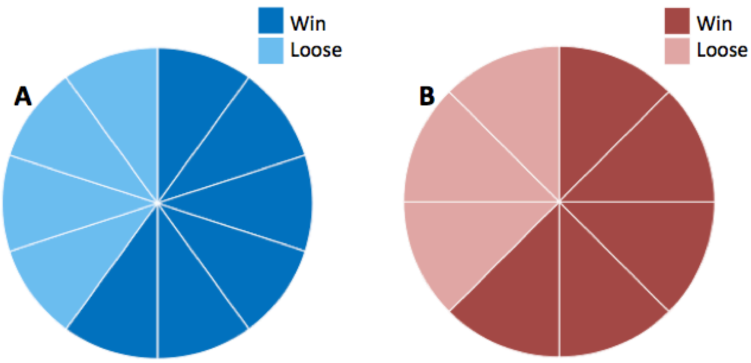


Figure 4-11 Operationalization of UNCERTAINTY - corresponding figure to item 2

In order to answer this item, test participants have to select the two answers from a multiple-choice drop-down menu[54] with 19 answer options. The menu is provided for the majority of the items of UNCERTAINTY and contains all correct answers for the whole item set of this dimension. Additionally, distractors have been added. They were constructed after analysing items' complexity considering which arithmetic errors are likely to be made and to which results they would lead. These results complement the drop-down menu. To choose the correct answer, test participants are asked to select the closest whole number from the drop-down list. The drop-down menu of the final DAC-test is presented in sub-section 9.8.6 of the appendix.

[54] It was decided to provide a multiple-choice drop-down menu instead of presenting an open text field for text/number entry, as the intention was to make the items as reliable as possible and maximise their scorability (Kline, 2015).

Comprehending the literal meaning of probabilistic information

Even though probabilities are present in our daily life, e.g. in package inserts of drugs or communicated on television weather forecasts, people do not have a common understanding of probabilities. Gigerenzer, Hertwig, van den Broek, Fasolo, and Katsikopoulos (2005) were able to show that even quantitative statements about probabilities are often misinterpreted. Based on this insight and their study, the third item of UNCERTAINTY intends to measure whether test participants comprehend the literal meaning of probabilistic information. It is operationalized by the question of what a probability given in the previous item means in that particular case. A corresponding multiple-choice drop-down menu with five potential explanations and an *I do not know* option are presented as a scale. The *I do not know* option was added, as it was done for several previous presented tasks because "[g]uessing is a major difficulty in tests, especially of ability and aptitude. The multiple-choice items where the distractors are equally good reduces the positive effects of guessing to a one-in-five chance, compared with the 50 per cent chance of true-false items" (Kline, 2015, p. 40).

Calculating a two-stage random experiment by calculating one conditional probability

The fourth item of this DAC-dimension assesses test participants' competence to handle a two-stage random experiment by calculating one conditional probability. In two to three sentences the item describes two events - A and B – and their probabilities, of which the second event depends on the first. Test participants are asked about the probability of event B after having experienced event A. For answering this item test participants have to select the answer from the 19-option drop-down menu. It is identical to the drop-down menu that is also presented for the second item of this DAC-dimension.

Calculating a random experiment with stochastically independent events

The fifth item of this DAC-dimension captures whether test participants are able to calculate a random experiment with stochastically independent events. Therefore, it presents two independent events and their probabilities. To answer this item, test participants can select the required answer from the 19-option drop-down menu.

Calculating expected values of two potential alternatives

The sixth item of UNCERTAINTY presents a short case with two alternatives and their probabilities of occurrence in order to operationalize whether decision-makers are able to calculate expected values of two potential alternatives. Test participants are asked to calculate the expected value for each alternative and write their answers without rounding up or down following the corresponding scale explanation (cf. Figure 4-12). This scale explanation is given to induce all test participants to use the same format, and to make the data analysis easier.

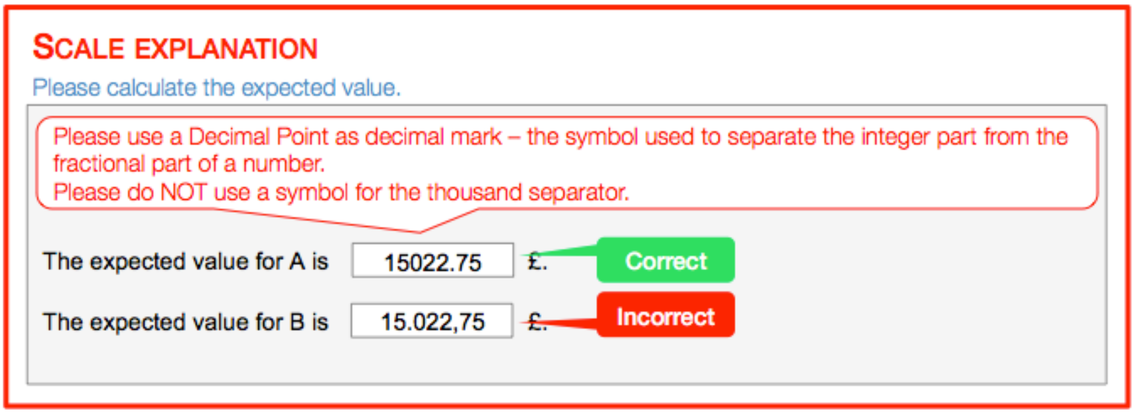


Figure 4-12 Operationalization of UNCERTAINTY – scale explanation

Calculating sensitivity/specificity with percentages and with natural frequencies

Items seven and eight request the calculation of sensitivity[55] or specificity[56] for a test presented in a short case of seven sentences providing total and conditional probabilities. While the seventh item of this DAC-dimension provides percentages of total and conditional probabilities, the eighth item presents its information in natural frequencies. Test participants are requested to select their answers again from the given 19-answer drop-down menu.

Scoring

For all items of UNCERTAINTY binary data is calculated for the process of scoring. Thus, each correct answer scores one point per item; but the second item, which asks for two answers, scores two points maximum. Thus, for the DAC pre-testing a maximum score of nine points could be reached for UNCERTAINTY.

4.1.7 Items of the Ability to Integrate Information

The seventh dimension of the DAC-test is the *ability to integrate information* (henceforth: INFORMATION). It describes the skill of subjects to combine assessments of

[55] Sensitivity is the true-positive rate, measuring the ratio of positives that are correctly identified as such by a test.

[56] Specificity is the true-negative rate, measuring the ratio of negatives that are correctly identified as such by a test.

uncertainties and values coherently, compare various alternatives and objectives, and correspondingly rank those alternatives in relation to their relevance for the decision-maker (e.g. Bazerman & Moore, 2009; W. Edwards, 1954; Goodwin & Wright, 2014; Keelin et al., 2009; Raiffa, 1968).

Test concept and theoretical basis

"Competent decision making requires several key skills including the ability to understand information, integrate information in an internally consistent manner, identify the relevance of information in a decision process, and inhibit impulsive responding." (Finucane & Gullion, 2010) Sound decision-making also demands the evaluation of objectives and uncertainties in a coherent manner (W. Edwards, 1954; Raiffa, 1968). So, the DAC-dimension INFORMATION captures the comparison of alternatives on the basis of individual objectives and the consideration of occurring uncertainties. Three steps of thought have to be executed; *comprehension, dimension weighting*, *probability consideration*.

Operationalization and layout

One decision-analytical approach to measure aspects of INFORMATION is the work of Finucane and Gullion (2010), under which the authors developed a tool to measure DMC of older adults in the context of health, finance, and nutrition decisions. Their test contains four different decision task types: comprehension, consistency, dimension weighting, and cognitive reflection. Out of those four measures, the *comprehension* measure (cf. Finucane, Mertz, Slovic, & Schmidt, 2005) and the *dimension weighting* measure (Finucane & Gullion, 2010) served as a construction guideline for operationalizing INFORMATION in the present research. By adapting the decision tasks of Finucane and Gullion (2010) to the context of the DAC-test and creating an additional item type for *probability consideration*, it is intended to capture the ability to compare alternatives on various objectives and given uncertainties, thus choosing the alternative that promises the highest chance to achieve the desired outcome (e.g. Bazerman & Moore, 2009; W. Edwards, 1954; Raiffa, 1968).

One secondary skill that is essential to successfully assess alternatives on various objectives is the skill to transform units and scales. As different objectives can vary in terms of their corresponding unit, such as costs in different currencies, decision-makers have to be able to transform scales to ease the comparison of alternatives. Consequently, a fourth item type was created to operationalize this secondary skill.

The process of choosing the best alternative is operationalized deterministically step-by-step in the DAC-test. However, before the four items for *comprehension, dimension weighting*, *probability consideration,* and *unit transformation* were presented, a fifth item was set first. Table 4-3 gives an overview of the five item types for INFORMATION. This first item requires the three steps of thought - *comprehension, dimension weighting*, and *probability consideration* - all at once, while items two, three, and four divide these steps so that test participants are led through this process of thinking. The idea behind that item-construction decision was that test participants might experience problems performing all steps of thought at once and thus not score at all. This circumstance would raise the question of which step(s) might have caused the problem and thereby which steps of thought could not be executed. Hence, by disconnecting these thinking steps it can be detected which exact step(s) might be the greatest challenge for test participants within this DAC-dimension. Additionally, the fifth item asks test participants to transform one objective unit into the unit of another objective.

Table 4-3 Overview of INFORMATION items of the DAC-test (pre-testing)

Object of measurement	Points to be reached
Comprehension, dimension weighting, probability consideration	3
Comprehension	3
Dimensions weighting	3
Probability consideration	3
Unit transformation	3

For this dimension of the DAC-test a single-item-per-page layout was chosen. The items of the final DAC-test are presented in sub-section 9.8.7 of the appendix.

The operationalization of INFORMATION starts with the presentation of a short decision case, which is connected to the main case of the DAC-test. It presents three individual objectives (1, 2, and 3) and three alternatives (A, B, and C) of the character. A table (cf. Table 4-4) containing the character's assessment of how well the three alternatives fit his/her objectives is presented and the expression of the fit visualised by symbols (here presented by *stars* - reaching from zero (worst fit) to five (best fit) stars).

Table 4-4 Operationalization of INFORMATION *– exemplarily assessment of alternatives*

	Alternative A	Alternative B	Alternative C
Objective 1		★★★	★★★★★
Objective 2	★★★★★	★★	★★★★
Objective 3	★★★	★★★★★	★★

Furthermore, test participants are informed that the character realises that his/her objectives weigh differently for each of the three alternatives. The weights of the three objectives are presented in percentages, adding up to 100%. In addition, it is announced that the presented decision situation contains one uncertainty. The first item is as follows:

Please rank the alternatives given the character's objectives and the prevalent risks. Start with the most preferred alternative.

Three drop-down menus, starting with the drop-down menu for the most preferred alternative and ending with the one for the least preferred alternative, are presented vertically. They provide the following answer options: "A; B; C; A and B; A and C; B and C; A, B, and C; /". To clarify, the following scale explanation (cf. Figure 4-13) is given.

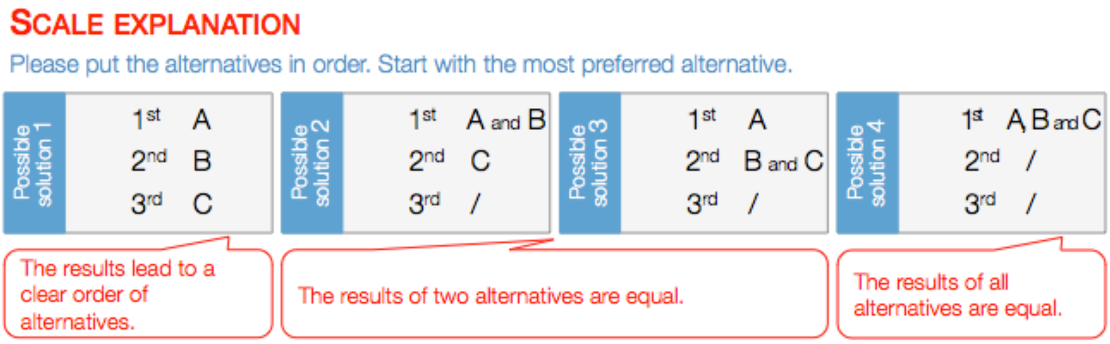


Figure 4-13 Operationalization of INFORMATION *– Scale explanation – Item 1*

As described above for the next three items of INFORMATION, the thinking steps of item one are further divided. Thus, on the next page the second item capturing comprehension presents a decision situation with three objectives and three alternatives, and a table (cf. Table 4-4) that shows the character's assessment of how well the alternatives fit his/her objectives. On the third page of this DAC-dimension, the third item capturing dimensions weighting presents how the character weighs the importance of his/her objectives. Within the fourth item capturing probability consideration it is explained that the character perceives an uncertainty for one of the alternatives.

INFORMATION items two, three, and four are:

Please put the alternatives in order according to the character's objectives for his/her decision. Start with the most preferred alternative.

Please put the alternatives in order according to the character's assessment of the relevance of his/her objectives. Start with the most preferred alternative.

Please put the alternatives in order according to the probabilities of the character's alternatives. Start with the most preferred alternative.

For answering the item, repeatedly the three drop-down menus and the corresponding scale explanation (cf. Figure 4-13) are presented.

The fifth item intends to measure whether test participants are able to transform the unit of a decision objective into another unit. Therefore, the item requires the calculation of a monetary scale into a 0 to 100 scale. The three alternatives from the case description of items one to four are presented again - this time with the expected costs for each alternative. Test participants are asked:

Please translate the prices into whole numbers between 0 and 100, given that

- £X is the maximum the character could spend,
- £Y is the minimum the character wants to spend,
- the more money saved the better, and
- the scale is interval-scaled, which means that the distance between 0 and 20 is equivalent to the distance between 60 and 80.

To answer the item, test participants are required to round to whole numbers and enter the correct number into three open text fields – one for each alternative.

Scoring

For the scoring, binary data is computed. Each correctly ordered alternative in the first, second, third and fourth item scores one point, which means three points maximum per item. Item five, for which three answers are required, scores one point per correct answer. Thus for the DAC pre-testing, a maximum score of 15 points could be attained for INFORMATION.

4.1.8 Items of the Ability to Plan to Implement a Decision

The *ability to plan to implement a decision* (henceforth: IMPLEMENTATION) is the eighth DAC-dimension. It describes the skill of subjects to set up a strategy for how to translate a cognitive decision into action by formulating concrete next steps and thinking ahead about possible obstacles and ways to address these obstacles (e.g. Hammond et al., 1999; Keelin et al., 2009; Mellers & Locke, 2007).

Test concept and theoretical basis

After having systematically compared the alternatives and reached clarity about which alternative to choose (Howard, 2007), decision-makers have to commit themselves to purposefully execute the made decision (Keelin et al., 2009). Thus, the motivational and volitional decision-making phases merge (Heckhausen & Gollwitzer, 1987). As the present research intends to assess DAC for complex decision problems, the transfer of intention, the motivational aspect of decision-making, into action, the volitional aspect of decision-making, is not possible to be operationalized plausibly on a quantitative level. To plan to implement a complex decision comprises an immense number of considerations and strategies and is hardly feasible in such a limited timeframe as the present DAC-test. Consequently, the aim for the sub-scale of IMPLEMENTATION was to operationalize only the motivational aspect of decision-making, that is the formulation of concrete next steps, the consideration of possible obstacles and the determination of ways to address these obstacles.

One approach to measure IMPLEMENTATION is the work of Lynch, Netemeyer, Spiller, and Zammit (2010), who developed the Propensity to Plan Scale - a self-assessment tool capturing the tendency of people to plan short- and long-term for time and money. However, two linked reasons led to the fact that their scale was neither applicable nor adaptable for the present research. First, Lynch et al.'s tool is a questionnaire, i.e. not capable of measuring actual performance like the DAC-test intends to do. Second, the tool thereby asked for the assessment of behaviour. The DAC-test, in contrast, aims to measure cognitive processes before the actual behaviour.

Operationalization and layout

According to the preceding DAC-dimensions, for IMPLEMENTATION, objective and quantitative items had to be found. As the literature does not provide an existing measure, it became clear that this DAC-dimension needed to be analysed exploratively first. Thus, the idea was to collect qualitative data in a small pilot study with 31 students and subsequently analyse and cluster their answers to define quantitative standards that could be set as a target value to achieve in a test.

Subsequently, the qualitative operationalization of IMPLEMENTATION was geared to the research on *action planning* and *coping planning* (cf. sub-section 2.4.3). In this framework it was intended to operationalize three facets of action planning and coping planning: (1) when-where-how plans (e.g. Gollwitzer, 1999) including the (2) anticipation of potential difficulties and (3) planning how to overcome those obstacles (e.g. Keelin et al., 2009; Sniehotta et al., 2006) when intending to translate a cognitive decision into action. For these three facets three different items were created, each of them presented on a single page: (1) formulating a commitment strategy with next steps as well as needed resources, (2) perceiving potential barriers and (3) formulating strategies to prevent or overcome them. The first item sets in at the point where the decision has already been made.

The qualitative items were presented with an open answer style in order not to imply that there is one correct answer. So, pilot study test participants were asked:

(1) Imagine you have just decided to <XXX[57]>. What would your plan of action look like in the next days, weeks and months to reach your goal(s)? Please describe for each step what you need to do. Which resources (e.g. help, additives) do you need to proceed with those steps?

(2) For each step, which difficulties could arise in transforming your plan into action?

(3) How would you try to prevent or overcome these difficulties?

The analysis of the qualitative results of the pilot study did not show a systematic pattern, to which the results could have been clustered. So it was impossible to formulate defined categories that could have been used as objective values for the DAC-test. As a result IMPLEMENTATION was not successfully operationalized for the pre-testing of the DAC-test.

Scoring

As explained above, the qualitative results of the pilot study did not allow for clustering the given answers since their variation was too complex. Thus, IMPLEMENTATION could not be operationalized for the DAC-test.

[57] Test participants of the pilot study were two decision scenarios presented; one referring to the INVEST case and one to the EDU case.

As described in this section, out of eight theoretical cognitive DAC-dimensions seven dimensions were operationalized by 19 different item types and by 69 items for the first version of the DAC-test.

The following section and its sub-sections 4.2.2 - 4.2.7 present the results of the pre-testing and corresponding item and distributional analyses of the DAC-prototype.

4.2 Item Analysis & Item Selection

In order to create a valid and reliable psychometric performance test, the first version of the two parallel DAC-tests (INVEST case and EDU case) with their 19 various item types (cf. section 4.1 and corresponding sub-sections) had to be pre-tested. As described in sub-section 3.2.3, qualitative and quantitative pre-tests were carried out in order to *revise the test*, *review the test score distribution*, and *inspect the test procedure* (Lienert & Raatz, 1998).

While sub-section 4.2.1 exposes the arguments for choosing one of the two cases for the main testing and thus, here for the item analysis and item selection, sub-section 4.2.2 gives an overview of the socio-demographic statistics of the quantitative pre-testing. Sub-section 4.2.3 presents the results of the item analysis and selection (revising the test) and sub-section 4.2.4 shows the results of the dimensionality analysis and the test score distribution (reviewing the test score distribution). Before sub-section 4.2.6 concludes the changes to the testing procedure for the main testing (inspecting the test procedure), sub-section 4.2.5 presents interpretations of the findings of the pre-testing. Sub-section 4.2.7 consolidates the modifications made to the pre-version of the DAC-test and explains the final-version of the test.

4.2.1 Data Set Selection

As presented at the beginning of chapter 4, two parallel cases of equal structure (cf. sub-sections 4.1.1 - 4.1.7) were contrived and tested in the pre-testing. Beforehand, both cases were pilot-tested qualitatively (cf. sub-section 3.2.3) and sporadically quantitatively by four test participants, who volunteered to run the last pilot-tests before the pre-testing. As a result, both cases were considered appropriate in terms of expenditure of time and being in a reasonable range.

However, the collected times for completing the INVEST case and the EDU case in the pre-testing and test participants' qualitative feedback[58] gave reason to select only one of the two cases for the main testing. On average, it took test participants of the pre-testing 158 minutes, i.e. 2 hours 38 minutes (*SD* = 58 minutes) to complete both the INVEST and the EDU case. The mean for processing the two cases and answering a first set of items of the criteria of validation (cf. sub-section 5.1), was 178 minutes, i.e. 2 hours 58 minutes (*SD* = 69 minutes).

The length of the test and the verbal feedback of test participants on decreasing concentration and motivation towards the second half of the test, led to the decision to reduce the test volume. Consequently, one case had to be neglected and the other selected for the main testing.

Even though the content of both cases was selected and designed to be as comparable as possible within the sample group, the EDU case was perceived as more generic and more accessible by the test participants[59]. So, the EDU case was chosen for the main testing. That is why the following analyses and considerations refer exclusively to the data set of the EDU case of the pre-testing.

4.2.2 Socio-Demographic Statistics of the Pre-Testing

For the pre-testing, the majority of the sample group (cf. corresponding to the criteria for the target group in sub-section 3.3.1) was assembled by three institutions in 2014: a group of the London School of Economics and Political Science summer school students taking the course of "Judgement and Decision Making for Management", voluntary students of the psychological department at Free University of Berlin, and voluntary students of the management department of the University of Bayreuth. Overall 196 people participated in the DAC pre-testing, of which 143 data sets could be employed for analyses. Three selecting criteria caused this reduction of 27.0% of the total sample: (1) according to the chosen target group for the present research (cf. sub-section 3.3.1), participants younger than 18 or older than 30 years, (2) participants with English skills that were self-rated lower than "good", and (3) test processing times shorter than 30 minutes

[58] Qualitative feedback was given mainly verbally by test participants who know the test constructor personally.

[59] After reviewing the test completion times and realising that due to the length of the two cases and the set of validation criteria only one case could be chosen for the main testing, an email was sent out to the 79 test participants whose email addresses were known. In that email test participants were asked: "Which of the two cases – INVEST case or EDU case – did you perceive as more generic and more accessible?" Forty-two persons replied, of which 31 selected the EDU case as the more generic and accessible case.

and longer than 3 hours for the EDU case. Participants who met one or more of these criteria were excluded from the analyses.

Table 4-5 gives an overview of the distribution of test participants according to their universities. Twenty-one test participants (14.7%) came from University of Bayreuth, 39 (27.3%) from Free University of Berlin, 58 (40.6%) people from London School of Economics, and 25 (17.5%) from multiple other sources such as former students of the Free University of Berlin who had heard about the test or colleagues and friends of the research group.

Table 4-5 Distribution by source of acquisition of the pre-testing sample

Source	Frequency	%
UB	21	14.7
Misc.	25	17.5
FU	39	27.3
LSE	58	40.6
Total	143	100.0

Note. UB = University of Bayreuth; Misc. = miscellaneous; FU = Free University of Berlin; LSE = London School of Economics and Political Science.

The mean age of the sample group was 25 years (*M* = 24.84; *SD* = 3.35; Range = 18-30). Fifty-four per cent of the 143 test participants assessed their English skills as “mother tongue” or “very good/business fluent” (cf. Table 4-6). The remaining 45.5% quoted their language skills as “good/conversant” or “fluent”.

Table 4-6 Frequencies of self-assessed English skills of the pre-testing sample

English level	Frequency	%
Good/conversant	18	12.6
Fluent	47	32.9
Very good/business fluent	19	13.3
Mother tongue	59	41.3
Total	143	100.0

Of the 143 test participants, 109 (76.2%) specified their gender: fifty-five of those test participants (50.5%) were female and 54 test participants (49.5%) were male.

The greatest national group were Germans with 45.5%, followed by Americans with 21.4%, Indians with 14.3%, and British with 10.7% of the valid data (112 test participants, cf. Table 4-7).

Table 4-7 Distribution by nation (pre-testing)

Nationality	Frequency	%	Valid %
German	51	35.7	45.5
American	24	16.8	21.4
Indian	16	11.2	14.3
British	12	8.4	10.7
Chinese	2	1.4	1.8
Scottish	2	1.4	1.8
Canadian	1	0.7	0.9
Hungarian	1	0.7	0.9
Italian	1	0.7	0.9
Polish	1	0.7	0.9
Vietnamese	1	0.7	0.9
NA	31	21.7	
Total	143	100.0	

Note. NA = missing data/not available.

The average processing time for taking the DAC-prototype including responding to the criteria for validation was one hour and 28 minutes (cf. Table 4-8). For only the EDU case, test participants of the pre-testing took on average one hour and one minute.

Table 4-8 Processing times in minutes of the pre-testing (EDU case & validation criteria)

Processing time ...	*N*	*Min*	*Max*	*M*	*SD*	*Mdn*	*IQR*
... for the EDU case	143	30	122	61.45	40.55	58.12	54.10 – 87.56
... for the EDU case and criteria of validity	143	38	148	88.45	62.28	85.79	80.49 – 137.43

Note. EDU case = Education case version of the DAC-test.

Table 4-9 shows how the sample group stated their educational level: Out of the 108 test participants who specified this information, almost 29.6% (32) have a university-entrance diploma and 48.1% (52) hold a Diploma, Bachelor’s, or Master’s degree.

Table 4-9 Distribution by degree of education of the pre-testing sample

Educational certificate	Frequency	%	Valid %
No school leaving certificate	1	0.7	0.9
Certificate after nine or 10 years of school education	18	12.6	16.7
University-entrance diploma	32	22.4	29.6
Certificate for apprenticeship	5	3.5	4.6
Bachelor's degree	44	30.8	40.7
Diploma or Master's degree	8	5.6	7.4
NA	35	24.5	
Total	143	100.0	

Note. NA = missing data/not available.

To the questions "What is your former, current or intended profession?" and "Which field does it belong to?" the pre-testing sample group made the following statements (cf. Table 4-10): By far the two largest groups of test participants came from Life, Physical, and Social Science Occupations with 41.3% (59), and from Business and Financial Operations Occupations with 23.8% (34).

Table 4-10 Distribution by former, current or intended profession of the pre-testing sample

Profession	Frequency	%
Arts, Design, Entertainment, Sports, and Media Occupations	5	3.5
Business and Financial Operations Occupations	34	23.8
Community and Social Services Occupations	1	0.7
Computer and Mathematical Occupations	10	7.0
Education, Training, and Library Occupations	3	2.1
Healthcare Support Occupations	9	6.3
Installation, Maintenance, and Repair Occupations	1	0.7
Life, Physical, and Social Science Occupations	59	41.3
Management Occupations	5	3.5
Military Specific Occupations	2	1.4
Office and Administrative Support Occupations	5	3.5
Transportation and Material Moving Occupations	9	6.3
Total	143	100.0

4.2.3 Results of Item Analysis & Consequences for the Item Selection

The present sub-section reveals how the constructed DAC-prototype performs empirically. Hence, it presents the statistical properties on each item and on the dimensional level. As mentioned in sub-section 3.2.3 the *item analysis* is intended to enhance reliability of the DAC-prototype by revising items and documenting the test point distribution. Consequentially, after the phase of item analysis, those items should be chosen that psychometrically fit the measurement of DAC best.

After omitting respondents listwise from the data analyses, who fell below the minimum processing time for completing the DAC-test, single NA responses were treated as wrong answers and therefore coded with “0”. This was done, as it was not possible to skip items in the DAC-test. NA responses were generated when test participants selected “I do not know” for some of the items with a multiple-choice response set. As in those cases test participants stated that they did not know the correct answer, “I do not know” was treated as wrong and so coded with “0” in the process of binary coding.

The following two criteria were considered for selecting the items for the final version of the DAC-test: First, in the context of the following analyses the *item discriminatory power coefficient* (r_{it}, cf. sub-section 3.2.3) measures how well an item fits the corresponding DAC-dimension. It was intended to select items that show a $r_{it} > .2$ (cf. Everitt, 2006; Fisseni, 2004). Second, the *item difficulty index* (P_i) explains the frequency with which an item was answered correctly by the sample group. It gives information on how difficult it is to solve an item. Based on Kline (2015) and Lienert and Raatz (1998) items with a p_i[60] between .8 and .2 were selected.

Most of the DAC-test items go along with a multiple-choice answer set, varying between two and 19 answer options. So for the majority of designed the items the influence of chance could not be precluded and, associated therewith, the corresponding probabilities of guessing right vary between 6% and 50%. For those items a formula to correct guessing – the relative difficulty ratio of Frisbie (1981) - was applied (cf. Lienert & Raatz, 1998; Moosbrugger & Kelava, 2012). In comparison to the regular formula for calculating item difficulty[61], which divides the number of correct answers by the number of

[60] $p_i = P_i / 100$

[61] According to Mossbrugger and Kelava (2012) the formula for calculating item difficulty in power tests is $p_i = n_C / N$.

participants, the corrected formula[62] (cf. Formula 4—1) subtracts the quotient of the number of incorrect answers and the number of answer options less one from the number of correct answers. This formula for the relative difficulty ratio was applied in place of the regular formula for each multiple-choice item of the DAC-test.

$p_{i\_c} = \frac{n_{c}-\left(\frac{n_I}{k-1}\right)}{N}$	p_{i_c} = corrected item difficulty index (P_i)/100 n_C = number of correct answers n_I = number of incorrect answers k = number of answer options N = number of participants

Formula 4—1 Calculation of the relative difficulty ratio (cf. Frisbie, 1981)

Theoretically, items with medium item difficulty of p_i = .5 differentiate best between test participants with a low and test participants with a high level of the characteristic value (Moosbrugger & Kelava, 2012, p. 87). However, for the present research a range of difficulty p_i between .2 and .8 appears to be appropriate as it is intended to assess people with low DAC as well as people with medium or high DAC (e.g. Kline, 2015, p. 143). As the current research does not aim to examine the extreme groups, i.e. people with either very low or very high characteristic value, items with p_i < .2 and p_i > .8 are excluded.

On the basis of these two essential statistics, and according to Kline (2015), all items meeting both criteria, discriminatory power coefficient and item difficulty index, were selected for the final test. Items that failed one or both of the criteria were examined for their relation to particular characteristics that could have caused this non-fit. Furthermore, all aspects of each DAC-dimension were covered content-wise and all DAC-dimensions provided at least eight items to enable a good comparability of scores among examinees.

The following paragraphs give an overview of the resulting statistical properties for each of the seven measured DAC-dimensions separately. Additionally, the reliability/internal consistency of each DAC-dimension is specified by Cronbach's alpha (α)[63]. The corresponding tables present the means (*M*), standard deviations (*SD*), item difficulty indices (p_i and p_{i_c}), discriminatory power coefficients (r_{it}), and whether the item

[62] It is known that this kind of guessing correction occasionally leads to negative item difficulty values, which cannot be interpreted. However, negative values can be treated as indicators for very difficult items that seem to be easier, i.e. such items encourage incorrect answers (cf. Lienert & Raatz, 1998, p.75; Moosbrugger & Kelava, 2012, p.80).

[63] Bland and Altman (1997) and Schmitt (1996) recommend an internal consistency coefficient (Cronbach's alpha) of α ≥.70 for newly constructed scales.

was selected for the main testing (MT) or not, including the reason for potential exclusion. All items of the DAC-test are coded so that higher values indicate better performance and therefore a higher decision competence.

Item analysis of the ability to envision one's objectives

As OBJECTIVES is operationalized by the item format of Bond et al. (2008), it does not consist of items in the "traditional way": The greatest difference in comparison to the other DAC-dimensions and their corresponding items is that in the case of OBJECTIVES the test participants determine the order and relevance of their objectives and thus the items individually. Hence, a comparison of all participants item by item is not feasible. According to Bond and colleagues (2008), the higher the percentage of self-generated objectives of the finally selected most relevant objectives, the better the test participants are able to envision their own objectives. In this sense, each self-generated objective of the final selected seven objectives scores one point.

Item analysis of the ability to recognise decision opportunities

Table 4-11 shows the item analysis results for the 12 items of OPPORTUNITIES in the pre-testing. As all items are multiple-choice items, the corrected item difficulty index p_{i_c} is calculated. The values for p_i are given for the sake of completeness. Out of the 12 items, nine items show good statistical properties with p_{i_c} between .2 and .8 and r_{it} above .2. Item DO_03 seems to be too easy as its p_{i_c} exceeds .8. As the p_{i_c} is negative and cannot be interpreted (cf. Footnote 62), p_i is taken into account. With p_i below .2, item DO_04 appears to be too difficult. The same circumstance applies to item DO_11, which also shows a negative p_{i_c}. Its p_i marginal meets the threshold of p_i >.2, but the discriminatory power coefficient does not fit and is too low with r_{it} = .11.

Table 4-11 Statistical properties of OPPORTUNITIES *(pre-testing)*

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	MT	Reason for potential exclusion
DO_01	.78	.42	.75	.55	.32	yes	
DO_02	.62	.49	.63	.68	.37	yes	
DO_03	.86	.35	.86	.89	.37		p_{i_c}> .8
DO_04	.12	.33	.12	-.30	.21		p_{i_c}< .2
DO_05	.90	.30	.90	.80	.26	yes	
DO_06	.22	.41	.36	.23	.47	yes	
DO_07	.73	.45	.68	.45	.68	yes	
DO_08	.62	.49	.51	.75	.63	yes	
DO_09	.59	.49	.58	.32	.68	yes	
DO_10	.86	.35	.83	.72	.32	yes	
DO_11	.22	.42	.22	-56	.11		r_{it} < .2
DO_12	.66	.48	.64	.31	.42	yes	

Note. *N* = 143; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; MT = selected for the main testing; DO = Item ID for the DAC-dimension of OPPORTUNITIES.

For the main testing, items DO_03, DO_04, and DO_11 were deleted. The remaining nine items were selected for the main testing. Cronbach's alpha for the nine remaining items of OPPORTUNITIES lies at α = .63.

Item analysis of the ability to assess decision fitness

The results of the item analysis for the eight items of FITNESS are displayed in Table 4-12. Two items show a negative corrected item difficulty index p_{i_c}. Hence, for these two items - DF_02 and DF_06 – the regular item difficulty index is taken into account. According to those values and the values for the discriminatory power index, all items fit the set thresholds for p_i/p_{i_c} and $r_{it.}$ Cronbach's alpha for the FITNESS is .33.

Table 4-12 Statistical properties of FITNESS *(pre-testing)*

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	MT	Reason for potential exclusion
DF_01	.80	.40	.80	.61	.26	yes	
DF_02	.45	.50	.45	-.09	.53	yes	
DF_03	.56	.50	.56	.21	.36	yes	
DF_04	.88	.32	.88	.76	.26	yes	
DF_05	.71	.46	.71	.41	.49	yes	
DF_06	.38	.49	.38	-.23	.53	yes	
DF_07	.64	.48	.64	.27	.36	yes	
DF_08	.71	.45	.71	.43	.49	yes	

Note. *N* = 143; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; MT = selected for the main testing; DF = Item ID for the DAC-dimension of FITNESS.

Item analysis of the ability to frame a decision

The items of FRAME are multiple-choice items as well; in this case with eight answer options. Therefore, the corrected item difficulty index p_{i_c} in comparison to the regular item difficulty index p_i is calculated. The statistical properties for the eight items of FRAME (cf. Table 4-13) all fulfil the target values for discriminatory power and item difficulty, even though the items FD_02 and FD_08 just meet them marginally.

Table 4-13 Statistical properties of FRAME *(pre-testing)*

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	MT	Reason for exclusion
FD_01	.47	.50	.47	.42	.36	yes	
FD_02	.24	.43	.24	.21	.36	yes	
FD_03	.58	.50	.58	.54	.70	yes	
FD_04	.42	.50	.42	.37	.45	yes	
FD_05	.64	.48	.64	.61	.66	yes	
FD_06	.64	.48	.64	.60	.57	yes	
FD_07	.41	.49	.41	.35	.55	yes	
FD_08	.21	.41	.27	.23	.28	yes	

Note. *N* = 143; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; MT = selected for the main testing; FD = Item ID for the DAC-dimension of FRAME.

Cronbach's alpha for this DAC-dimension is .51. The DAC-dimension of FRAME did not show an adequate Cronbach's alpha. However, the items of FRAME were kept for the main testing. One adaption that was undertaken for this scale was the presentation of a definition for the requested terms of the item (cf. sub-section 4.2.5).

Item analysis of the ability to identify relevant alternatives

Only one item out of the 10 items of ALTERNATIVES (cf. Table 4-14), IA_01, violates both target values of discriminatory power and item difficulty. The corrected item difficulty indices p_{i_c} for items IA_02 and IA_05 are negative and not interpretable. Thus, in both cases the regular item difficulty index p_i is examined. Consequently, item IA_02 appears to be too difficult.

Table 4-14 Statistical properties of ALTERNATIVES (pre-testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	MT	Reason for exclusion
IA_01	.92	.28	.92	.83	.16		$p_{i\_C} > .8$ & $r_{it} < .2$
IA_02	.47	.50	.12	-.05	.53		$p_{i\_C} < .2$
IA_03	.82	.39	.75	.64	.53	yes	
IA_04	.77	.42	.75	.54	.42	yes	
IA_05	.33	.47	.41	-.34	.21	yes	
IA_06	.66	.48	.75	.31	.58	yes	
IA_07	.67	.47	.71	.34	.63	yes	
IA_08	.83	.38	.73	.65	.47	yes	
IA_09	.74	.44	.71	.48	.37	yes	
IA_10	.61	.49	.59	.22	.53	yes	

Note. *N* = 143; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; MT = selected for the main testing; IA = Item ID for the DAC-dimension of ALTERNATIVES.

Out of the 10 ALTERNATIVES items the eight remaining were selected for the main testing. Unfortunately, this deletion led to a decrease in Cronbach's alpha to $\alpha = .44$.

Item analysis of the ability to deal with uncertainty

Table 4-15 gives an overview of the statistical properties of the 10 items of UNCERTAINTY. Apart from items DU_07 and DU_08, all items are multiple-choice with 19 answer options for which corrected item difficulty indices were calculated.

It becomes apparent that the four items – DU_07, DU_08, DU_09, and DU_10 – do not fit the two target values of discriminatory power r_{it} and item difficulty p_i/p_{i_c}. They seem to be too hard to solve.

Table 4-15 Statistical properties of UNCERTAINTY *(pre-testing)*

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	MT	Reason for exclusion
DU_01	.85	.36	.83	.69	.37	yes	
DU_02	.43	.50	.56	.39	.84	yes	
DU_03	.38	.49	.42	.35	.89	yes	
DU_04	.47	.50	.47	.34	.37	yes	
DU_05	.73	.44	.78	.72	.47	yes	
DU_06	.24	.43	.32	.20	.68	yes	
DU_07	.08	.28	.10	*	.16	yes	
DU_08	.03	.18	.05	*	.16	yes	
DU_09	.08	.28	.08	.03	.21		$p_{i\_c} < .2$ & $r_{it} < .2$
DU_10	.15	.36	.15	.11	.42		$p_{i\_c} < .2$ & $r_{it} < .2$

Note. *N* = 143; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; MT = selected for the main testing; DU = Item ID for the DAC-dimension of UNCERTAINTY.

* Items are not multiple-choice items. Guessing correction is not necessary to calculate item difficulty index.

Even though four items of UNCERTAINTY did not show good statistical properties in the item analysis, only the last two items – DU_09 and DU_10 – were excluded from the main testing version of the DAC-test. Items DU_07 and DU_08 were kept due to content reasons as they are referring to the calculation of expected values, which is one of the main tools in decision analysis (e.g. McNamee & Celona, 2005). For the main testing, a cue was introduced to ensure that all participants understood what was meant by *expected value* (cf. sub-section 4.2.5). For this DAC-dimension the deletion of two items resulted in a decrease of Cronbach's alpha to α = .64.

Item analysis of the ability to integrate information

The 15 items of INFORMATION are presented in Table 4-16 with the results of their item analysis. A particular characteristic of this dimension is that its items are arranged in triplets. Thus, items II_01, II_02, II_03 belong to the same instruction. In the same style items II_04, II_05, and II_06 are counted as one instruction. The same principle applies to the other nine items. Except for items II_13, II_14, and II_15, all items are multiple-choice items with eight potential answer options. For these items the corrected item difficulty index p_{i_c} was calculated. According to the statistical properties only items II_02 and II_03 seem to be too difficult with a corrected item difficulty index p_{i_c} of .13 and .11. All other items meet the threshold for both item difficulty and item discriminatory power.

Table 4-16 Statistical properties of INFORMATION *(pre-testing)*

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	MT	Reason for exclusion
II_01	.37	.49	.37	.28	.89		Test economy
II_02	.24	.43	.24	.13	.74		$p_{i\_C} < .2$, Test economy
II_03	.22	.42	.22	.11	.53		$p_{i\_C} < .2$, Test economy
II_04	.78	.42	.76	.74	.53	yes	
II_05	.64	.48	.61	.58	.68	yes	
II_06	.68	.47	.63	.63	.63	yes	
II_07	.80	.40	.81	.78	.42	yes	
II_08	.22	.42	.31	.21	.84	yes	
II_09	.19	.39	.24	.22	.68	yes	
II_10	.17	.38	.27	.24	.74	yes	
II_11	.24	.43	.34	.29	.63	yes	
II_12	.21	.41	.25	.20	.63	yes	
II_13	.28	.45	.31	*	.68	yes	
II_14	.26	.44	.29	*	.68	yes	
II_15	.27	.45	.29	*	.63	yes	

Note. $N = 143$; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; MT = selected for the main testing; II = Item ID for the DAC-dimension of INFORMATION.

* Items are not multiple-choice items. Guessing correction is not necessary to calculate item difficulty index.

Out of the 15 items of INFORMATION, 12 items were selected for the main testing. Even though item II_01 meets the thresholds of p_{i_c} and r_{it}, the other two items of the triplet - items II_01 and II_02 - seem to be relatively difficult to solve. In order to shorten the test and thereby reduce the workload for future test participants items II_01, II_02, and II_03 were deleted. Also, competence-wise the triplet is covered by the nine items II_04 - II_12 (cf. sub-section 4.1.7). In terms of its internal consistency the dimension of INFORMATION shows a very good Cronbach's alpha of $\alpha = .85$.

According to the literature of scale construction (e.g. Moosbrugger & Kelava, 2012), it is worthwhile building a scale that contains of items with different levels of item difficulty resulting from various requirements. Items with a middle difficulty index of p_i/p_{i_c} around .5 are best suited to differentiate between people with high and people with low level of the to-be-measured construct. However, for a test it is also important to contain items low in difficulty ($p_i/p_{i\_c} > .60$) so that no test participant ends up with zero points and items high in difficulty ($p_i/p_{i\_c} < .40$) and so that only a few test participants are able to solve all items. Table 4-17 gives an overview of how the item difficulty indices of each DAC-dimension of the pre-testing are distributed. The item difficulty indices for all dimensions, apart from UNCERTAINTY, range between .2 and .8. Since UNCERTAINTY consists of two very difficult items (DU_07: $p_i = .10$ and DU_08: $p_i = .05$), the average of the scale's difficulty is lower than the means of the other scales. OPPORTUNITIES appears to be the easiest scale with p_{i_c} = .53. However, all means of item difficulty indices range between .36 and .53.

Table 4-17 List of item difficulty indices (p_i /p_{i_c}) per DAC-dimension (pre-testing)

DAC-dimension	*M*	SD	*Range*
DO	.53	.20	.23 – .80
DF	.44	.17	.21 – .76
FD	.42	.15	.21 - .61
IA	.45	.15	.22 - .65
DU	.36	.23	.05 - .72
II	.40	.21	.20 - .78

Note. *N* = 143; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

Besides examining item difficulty and discriminatory power, the reliability of each DAC-dimension was measured. Therefore the internal consistency by Cronbach's alpha (Cronbach, 1951) was calculated, "which ... has become routine practice in virtually all psychological and social science research in which multiple-item measures of a construct are used" (Schmitt, 1996, p. 350). Interdisciplinary, a common presumption on the sufficient threshold of Cronbach's alpha exists. According to Bland and Altman (1997), Nunnally et al. (1978), or Schmitt (1996), a Cronbach's alpha of at least .70 is desirable for developing a new scale. Table 4-18 summarises the parameters of internal consistency for each DAC-dimension after the item selection.

At this point, the results of the analysis for internal consistency are only described and not interpreted or discussed. Further critical considerations regarding these values can be found in sub-section 4.2.5.

With a Cronbach's alpha of α = .33[64], FITNESS shows the lowest internal consistency, followed by ALTERNATIVES with α = .44, and FRAME with α = .51. The values for OPPORTUNITIES (α = .63) and UNCERTAINTY (α = .64) almost meet the desirable threshold for Cronbach's alpha. INFORMATION presents a good value of α = .85.

[64] Cronbach's alpha is sensitive to the number of items (e.g. Eid & Schmidt, 2014, Nunnally, 1978), especially for scales with fewer than 10 items. That could be one reason why, even though items with poor statistical properties have been excluded from this calculation due to the item analysis, Cronbach's alpha is still pretty low. «Theoretically, the larger the number of items in a scale, the more reliable will be the scale.» (R. A. Peterson, 1994, p. 389)

Table 4-18 Overview of internal consistencies of the pre-testing after item selection

DAC-dimension	*α*	Number of items
EO	-	-
DO	.63	9
DF	.33	8
FD	.51	8
IA	.44	8
DU	.64	8
II	.85	12

Note. DAC = Decision-analytical competence; *α* = Cronbach's alpha; EO = Ability to envision one's objectives; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

4.2.4 Test Score Distribution & Dimensionality of DAC

Having selected the items for the main testing, the item analysis also serves to test the distribution of test points, examining whether the test and its sub-scales are normally distributed. Table 4-19 gives an overview of the descriptive statistics for the calculated sum scores of the seven operationalized DAC-dimensions of the pre-testing.

The Kolmogorov-Smirnov test (Massey, 1951) and the more powerful Shapiro-Wilk test (Shapiro & Wilk, 1965) are known for being case sensitive (cf. Mason & Schuenemeyer, 1983), i.e. they are very sensitive to small variations given the large sample group. Thus, the values of skewness and kurtosis (cf. Table 4-19), and graphical distributions (cf. Figure 4-14 and Table 4-20) were analysed to examine the distribution of the present data.

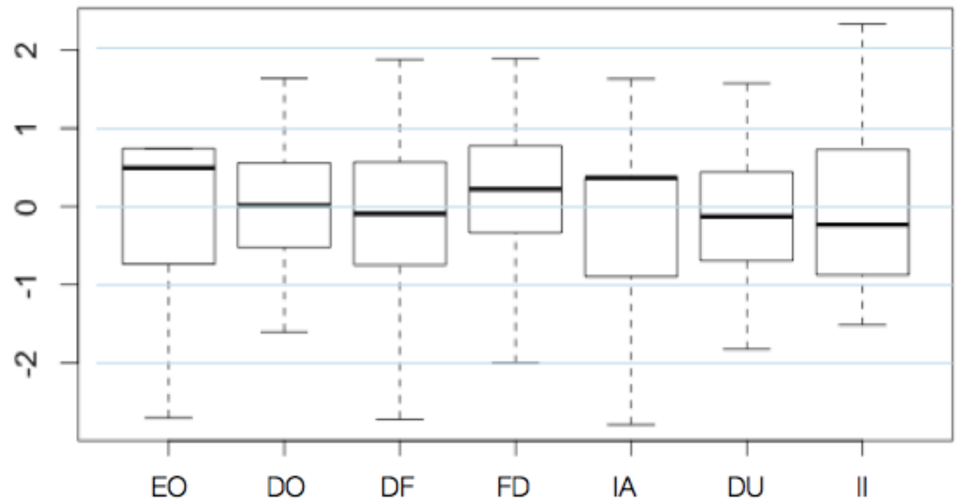
Table 4-19 Descriptive statistics of the sum scores of the seven operationalized DAC-dimensions

DAC-dimension	*M*	*SD*	Potential range	Observed range	Skew	Kurtosis
EO sum score	5.50	2.03	0-7	0-7	-1.31	0.65
DO sum score	5.97	1.85	0-9	0-9	-0.63	0.28
DF sum score	5.14	1.52	0-8	1-8	-0.25	-0.08
FD sum score	3.60	1.80	0-8	0-8	-0.29	-0.59
IA sum score	5.42	1.59	0-8	0-8	-0.59	0.32
DU sum score	3.22	1.77	0-8	0-8	0.32	-0.41
II sum score	4.73	3.12	0-12	0-12	0.68	-0.19

Note. N = 143; DAC = Decision-analytical competence; skew = skewness; EO = Ability to envision one's objectives; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

According to George and Mallery (2011) the values for skewness and kurtosis between -1 and +1 are considered very acceptable in order to prove the normal univariate distribution for psychometric uses. Even though all dimensions apart from OBJECTIVES show slightly asymmetrical distributions (skew ≠ 0), they are located within the acceptable range for normal distribution. The skewness of OBJECTIVES by far exceeds this range. It presents a negative value of -1.31 and so reveals that the majority of test participants of the pre-testing have reached high scores of this scale. While OPPORTUNITIES, FITNESS, ALTERNATIVE, and FRAME present a negative skew as well and thus a longer tail to the left, UNCERTAINTY and INFORMATION show a longer tail to the right, i.e. more than fifty per cent of test participants have gained less than half of the to-be-reached points of the two DAC-dimensions. According to the kurtosis values (kurtosis of a Gaussian distribution = 0) OBJECTIVES, OPPORTUNITIES, and ALTERNATIVES have a more peaked distribution and FITNESS, FRAME, UNCERTAINTY, and INFORMATION have flatter distributions than a *normal distribution*.

Figure 4-14 shows the scaled boxplots (*M* = 0; *SD* = 1; statistical outliers of two standard deviations and more are excluded) of the seven DAC-dimensions after the item selection. As the skewness values of OBJECTIVES already suggests, the boxplot of OBJECTIVES reinforces a strong asymmetry. The OBJECTIVES-boxplot shows a ceiling effect (Cramer & Howitt, 2004, p. 21, [entry "ceiling effect"]), which appears if "the majority of values obtained for a variable approach the upper limit of the scale" (Zedec, 2014, p. 38). Therefore, the variance within a variable cannot be measured precisely. In the case of OBJECTIVES, too many test participants have reached a high score of this DAC-dimension and thus, the item(s) do not discriminate well enough.



Note. EO = Ability to envision one's objectives; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information

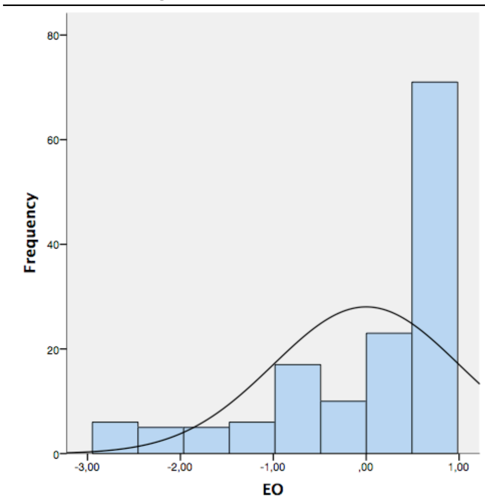
Figure 4-14 Boxplots of the seven scaled DAC-dimensions after item selection (pre-testing)

Table 4-20 presents the sum scores with overlaid curves of a normal distribution for each DAC-dimension, the overall score and thus, the DAC test point distribution. Also here, OBJECTIVES shows an extreme in the area of high scores. Looking at the data of OBJECTIVES, on average test participants have generated eight objectives (*M* = 7.95; *SD* = 5.06; Range: 0 – 30). From the master list they have recognised on average 20 objectives (*M* = 19.88; *SD* = 9.27; Range: 1 - 43) and they have matched on average 32.15% of the recognised objectives (*SD* = 22.05%; Range: 0% - 95.35%) with their self-generated objectives. Given these results, it seems as test participants might not put too much effort into the generation task. However, Bond and colleagues (2008) admit that a high percentage of self-generated objectives of the finally selected objectives might happen if test participants "overrate the importance of self-generated objectives" or "notice the substantial number of recognized objectives and thus overrate their importance" (p.61). Additionally, one adaption that has been made in comparison to the original measuring procedure of Bond et al. (2008, 2010) is that the number of objectives, which had to be chosen in the last step, was reduced from 10 to seven. This reduction could also have caused the ceiling effect, which does not allow variance to be adequately measured within the sample group and thus, differentiating between test participants. For this reason and to run further analyses on the assumption that the data of the pre-testing is normally distributed (cf. lowest plot of right column of Table 4-20), the data of OBJECTIVES had to be excluded from the following analyses.

Table 4-20 Histogram of each DAC-dimension and the overall DAC-test score (pre-testing)

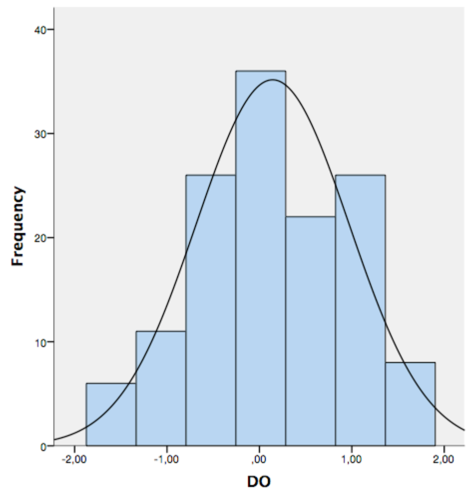
Histogram of scaled values

EO

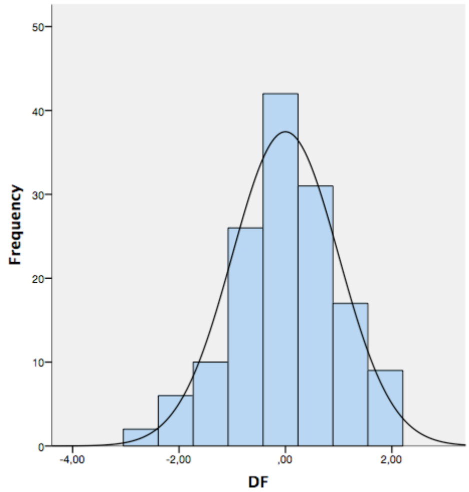


Histogram of scaled values

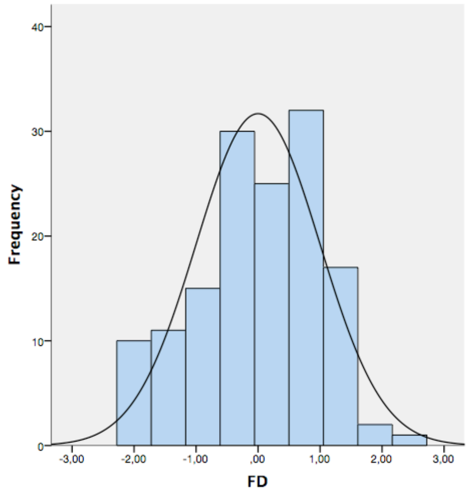
DO



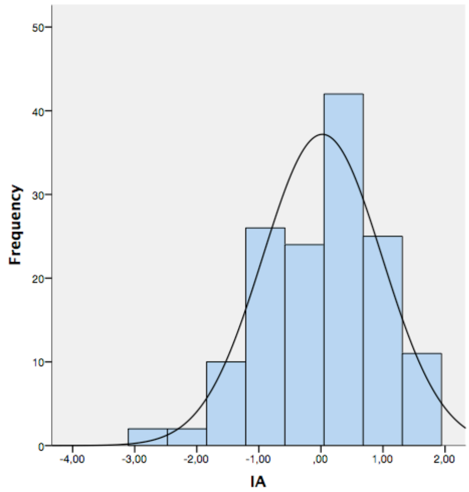
DF



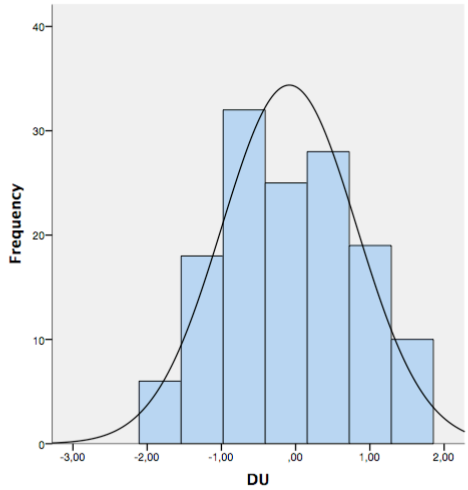
FD



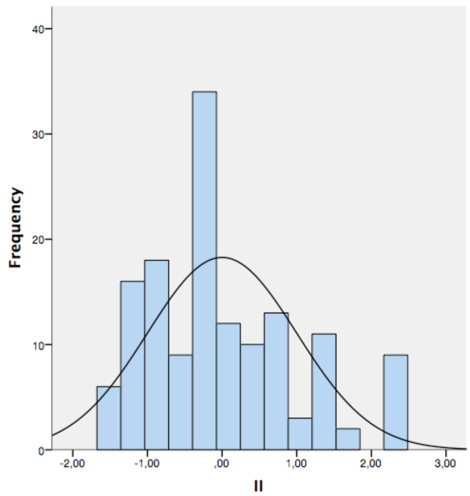
IA



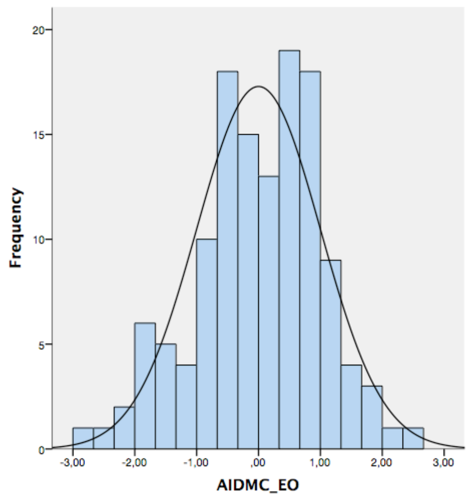
DU



II



DAC w/o EO



Note. DAC = Decision-analytical competence; EO = Ability to envision one's objectives (*N* = 143; *M* = -4.44E-16; *SD* = 1.00); DO = Ability to recognise decision opportunities (*N* = 135; *M* = 0.14; *SD* = 0.83); DF = Ability to assess decision fitness (*N* = 143; *M* = -1.40E-11; *SD* = 1.00); FD = Ability to frame a decision (*N* = 143; *M* = 8.39E-11; *SD* = 1.00); IA = Ability to identify relevant alternatives (*N* = 142; *M* = 0.2; *SD* = 0.96); DU = Ability to deal with uncertainty (*N* = 138; *M* = -0.09; *SD* = 0.91); II = Ability to integrate information (*N* = 143; *M* = 2.45E-10; *SD* = 1.00); DAC w/o EO = Decision-Analytical Competence without ability to envision one's objectives (*N* = 130; *M* = 1.01E-16; *SD* = 1.00).

[a] Statistical outliers of two standard deviations and more are excluded.

As the composition of the sample group was compatible and the display of the frequency distribution (cf. lowest plot of right column of Table 4-20) justifies the supposition of a normal distribution, it is assumed that the DAC-prototype scores are normally distributed.

Table 4-21 presents bivariate correlations among the six remaining DAC-dimensions. They are all positive and range from low[65] r_p = .21 (among ALTERNATIVES and FITNESS) to strong r_p = .53 (among INFORMATION and UNCERTAINTY) with a moderate overall mean correlation of r_p = .32.

Table 4-21 Pearson correlations (r_p) between DAC-test sub-scales

	DO	DF	FD	IA	DU	II
DO	1					
DF	0.27**	1				
FD	0.22**	0.43***	1			
IA	0.28***	0.21*	0.33***	1		
DU	0.30***	0.32***	0.34***	0.29***	1	
II	0.28***	0.33***	0.34***	0.29***	0.53***	1

Note. DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

[a] Mean correlation: r_p = .32.

*Two-sided; *$p < .05$; **$p < .01$; ***$p < .001$.

Another aspect examined for the DAC-prototype, is the dimensionality of the DAC-test by an exploratory factor analysis (EFA). An EFA with promax rotation[66] (factors are non-orthogonal and thus related; Gorusch, 1983, p. 203/204) was applied to the six DAC-dimensions in order to explore and reveal one or more potential underlying constructs. As a result, a one-factor model was obtained (cf. Figure 4-15).

[65] According to Cohen (1988) to evaluate correlations the following thresholds are set: low correlation → r ≥ .1, moderate correlation → r ≥ .3, strong correlation → r ≥ .5.

[66] "Any of several methods in factor analysis by which the researcher attempts to relate the calculated factors to theoretical entities. This is done differently depending upon whether the factors are believed to be correlated (oblique) or uncorrelated (orthogonal)." (Vogt, 1993, p. 91)

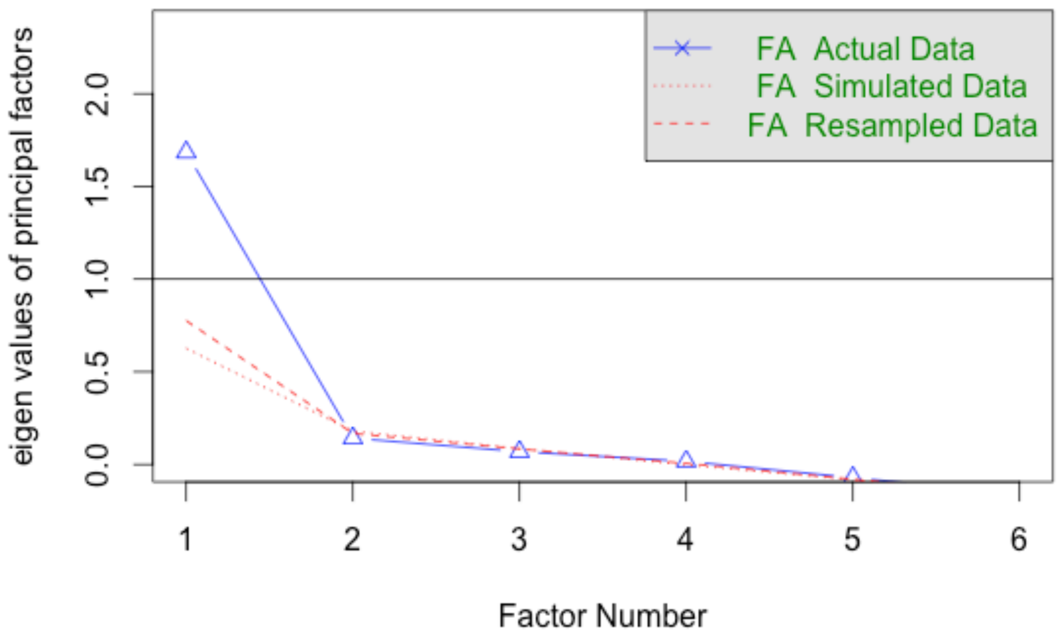


Figure 4-15 Parallel analysis screen plots for the data of the pre-testing

The one-factor solution explains 29% of the variance of the pre-testing data. All of its loadings meet the cut-off of $\lambda \geq .30$ (Hair, Tatham, Anderson, & Black, 1998, p. 112) and are presented in Table 4-22. Respectively to Tabachnick and Fidell's (2007) and Comrey and Lee's (1992) distinction of cut-offs for factor loadings, four of the six dimensions show good ($\lambda \geq .55$) or very good ($\lambda \geq .63$) fit. According to these results, it appears reasonable that an underlying general ability of DAC creates a moderate percentage of performance variance.

Table 4-22 Loadings for the EFA one-factor DAC model (pre-testing)

DAC-dimension	Loadings of one-factor model
II	.66
DU	.62
FD	.57
DF	.55
IA	.45
DO	.36
Eigenvalue	1.76
Variance explained	29%

Note. EFA = Exploratory factor analysis; DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

Cronbach's alpha of the six DAC-dimensions' sum scores accounted for $\alpha = .72$. Thus, the threshold of .70 for Cronbach's alpha (cf. Bland & Altman, 1997; Nunnally et al., 1978; Schmitt, 1996) was met in the pre-testing for the DAC-test, which is composed of the sum scores of the six DAC-dimensions.

4.2.5 Interpretations of the Findings of the Pre-Testing

As presented by Table 4-18, none of the DAC-dimensions except INFORMATION meets the desired threshold for Cronbach's alpha of .70. Before this sub-section discusses what that means for the five remaining dimensions, it presents potential reasons for these reliability results and outlines how to proceed, followed by a short discourse on the reliability coefficient and its minimally acceptable value for scale development.

It is rather surprising that the technical literature provides only a small number of recommendations on a sufficient threshold for Cronbach's alpha. However, these recommendations do have two things in common: "they indicate that the required degree of reliability is a function of the research purpose, whether the research is exploratory, applied, or so forth ... [and they all miss] an empirical basis, a theoretical justification, or an analytical rationale." (R. A. Peterson, 1994, p. 381). So, Nunnally (1967) postulates a Cronbach's alpha between .50 and .60 as being within the minimal acceptable range for preliminary research. Eleven years later, Nunnally, Bernstein, and ten Berge (1978) recommend a minimal threshold of .70 without giving an explanation for the increase of the targeted value. Now, the threshold of .70 is widely accepted in various research fields such as business, education, psychology (R. A. Peterson, 1994). Thus, it is also applied to the present research.

While Cronbach's alpha greater than .70 is taken as a checkpoint in the scale construction phase, lower values lead to further analyses. The received low Cronbach's alphas of the pre-testing entailed the following three considerations:

First, "[t]heoretically, the larger the number of items in a scale, the more reliable will be the scale." (R. A. Peterson, 1994, p. 389) Especially for scales with less than 10 items, Cronbach's alpha is sensitive to the number of items (e.g. Eid & Schmidt, 2014, Nunnally, 1978). That could be *one* reason why, even though items with poor statitistical properties have been excluded due to the item analysis, the five DAC-dimensions with fewer than 10 items - all but INFORMATION – end up with relatively low Cronbach's alphas.

Second, if a scale aims to capture a complex construct it is difficult to reach high values for Cronbach's alpha (Krüger, Parchmann, & Schecker, 2013). In comparison, according to Schmitts (1996) a low Cronbach's alpha does not have to be a reason for dropping a scale: "When a measure has other desirable properties, such as meaningful content coverage of some domain and reasonable unidimensionality, this low reliability [(Schmitt talks about a Cronbach's alpha of .49)] may not be a major impediment to its use". (p. 351/352) Thus, the underlying conceptual complexity of some DAC-dimensions, as for instance FRAME (cf. sub-section 4.1.4) could have led to low internal consistency.

Third, if a scale's reliability coefficient cannot be increased by deleting items based on the item analysis, the items' interrelatedness has to be examined by an EFA to investigate the scale's dimensionality (e.g. Schmitt, 1996).

The first consideration was present during the data analysis of the pre-testing. However, due to the test volume, which was to be kept as short as possible, raising the number of items for each of the five DAC-dimensions with medium or low Cronbach's alphas, was not an option; particularly not because this would be contrary to the classical meaning of analysing a scale's reliability by Cronbach's alpha.

For the second consideration, the five DAC-dimensions with medium or low Cronbach's alpha values were examined carefully by the research team – comparing its theoretical content and the methodological coverage of this content. All scales were considered as fitting. This conclusion represented the first doubts about the data quality of the pre-testing. The test volume with both cases and the reported decreasing motivation and concentration of some test participants (cf. sub-section 4.2.1) reinforced the suspicion that not all test participants gave their best to take the test, especially not in the second case – the EDU case (cf. sub-section 4.2.1). As those DAC-dimensions show the lowest alpha (FITNESS with α =. 33 and ALTERNATIVES with α =. 44), with the smallest answer option set, it has to be questioned whether some test participants might have just guessed.

According to the third consideration and the suspicion that the quality of data of the pre-testing might show weak points, it was decided to keep all DAC-prototype sub-scales for the main testing. By approaching a more homogeneuos sample group and setting financial incentives (cf. sub-section 3.3.3), an attempt was made to ensure data quality for the main testing. If the values for Cronbach's alpha turned out to be similar to the values of the pre-testing, then the dimensionality of the scales would be examined by confirmatory factor analyses (CFA). In that case, other measures for reliability have to be found (cf. Table 5-9, sub-sections 6.1.1, 6.1.2, and section 6.2).

4.2.6 Changes to the Test Procedure for the Main Testing

Besides the reduction of items with poor statistical properties (cf. sub-section 4.2.3), the pre-testing also helped to reveal potential sources of error, and thus to improve the measuring procedure. In this context, the participants in the main testing received the following instructions at the beginning of the test:

- Please make sure to follow all test instructions precisely.
- The test will take approx. 1.5 hours to complete. So please allow yourself enough time and answer each item with consideration and very thoroughly.
- If you take a break, please leave the browser window open. If you close it, you will have to start over again.
- You cannot go back in the test. If you press enter the system will guide you to the next page.

After the verbal feedback of colleagues and some test participants that displaying a graphical progress indicator would enable test participants to track their progress during the test and thereby support them, the question arose as to how helpful such a tool would be. As studies have shown that such an indicator tends to be effective only for short surveys (e.g. Couper, Traugott, & Lamias, 2001; Heerwegh & Loosveldt, 2006) and in "long surveys they may be more discouraging than encouraging" (Dillman et al., 2014, p. 326), the idea of implementing a graphical progress indicator was discarded.

According to the operationalization of the seven DAC-dimensions three adaptions were made:

OBJECTIVES: As a consequence of the observed ceiling effect concerning OBJECTIVES, the number of objectives to be selected in the last step of the item instruction (cf. sub-chapter 4.1.1) was increased from seven to 10 objectives.

FRAME: To ensure that all test participants are able to understand what is meant by the presented descriptions/terms of the instruction for FRAME (cf. sub-chapter 4.1.4), and thus grasp the descriptions' meaning in the decision-making context, term definitions were added. Figure 4-16 shows the definitions of the descriptions, which are provided by manoeuvring the cursor over the underlined comment *cf. definitions* at the end of the item instruction.

DEFINITIONS

- Objectives: a decision maker's individual preferences/values/aims
- Conflicting objectives: objectives that are at odds with other objectives of the decision maker
- Alternative(s): options the decision maker can choose from
- Uncertainty: judgment of the likelihood of different outcomes
- Time frame: available time for making the decision
- Decision maker: character who has to make a decision
- Resource: a source or supply to make a decision, typically of limited availability
- Friend: friend of the decision maker
- Family: family member of the decision maker
- Location: location where the decision takes place

Figure 4-16 Added definitions for decision-analytical terms requested for FRAME

UNCERTAINTY: In order to create fair initial conditions regarding solving the sixth item of UNCERTAINTY (cf. sub-section 4.1.6), a definition of *expected value* is presented to support those test participants who do not know the term from previous education or working experience. Figure 4-17 shows the definition which appears by moving the cursor over the underlined comment *cf. definitions* at the end of the item instruction.

DEFINITION

Expected value: Anticipated value for a given investment. In statistics and probability analysis, expected value is calculated by multiplying each of the possible outcomes by the likelihood that each outcome will occur, and summing all of those values. By calculating expected values, decision maker can choose the scenario that is most likely to give them their desired outcome.

Figure 4-17 Added definition for expected values requested for the sixth item of UNCERTAINTY

4.2.7 The DAC-Test Version for the Main Testing

Once the item and distributional analyses were concluded, the final version of the DAC-test was assembled. It contains 16 various item types for operationalizing seven different dimensions of DAC. Hence, three item types have been deleted (two items of UNCERTAINTY and one item of INFORMATION). The overall item number was cut from 70 to 63 items, as items with poor statistical properties ($p_i < .2$ and/or $r_{it} < .2$) were excluded as well. The calculated parameter for internal consistency, Cronbach's alpha, ranged from $\alpha = .33$ (poor fit) to $\alpha = .85$ (good fit). However, Cronbach's alpha of the complete DAC-scale (without taking into account OBJECTIVES) ended up at $\alpha = .72$, which met the goal for creating a new measure (e.g. Bland & Altman, 1997).

According to the values of skewness and kurtosis and the graphical display of the dimensional distributions and the overall DAC-score distribution, a normal distribution of the DAC-prototype for the present data is assumed. The results of a Pearson correlation present a moderate inter-dimensional correlation of r_P = .32 and the output of an EFA shows a one-factor solution, which explains 29% of the variance.

Chapter 5 is devoted to the analysis of goodness criteria for the final DAC-test. While sub-section 5.1 presents the chosen criteria of validation including the selected scales, sub-section 5.2 shows the results of the *empirical control* (cf. sub-section 3.2.4).

5 ANALYSING GOODNESS CRITERIA OF THE DAC-TEST

The pre-testing served to revise the prototype of the DAC-test, to review its test score distribution and dimensionality, and to inspect the testing procedure. The resulting final version of the DAC-test was prepared for the main testing to analyse the test's quality criteria.

Section 5.1 gives an overview of the selected criteria for validation, which have been applied to examine the DAC-test's validity. Therefore, sub-sections 5.1.1 to 5.1.6 introduce six psychological constructs. In sub-section 5.1.7 the obtained socio-demographic variables are presented and, for the sake of completeness, a short summary of the pre-tested and rejected criteria for validation is given in sub-section 5.1.8.

Section 5.2 presents the results of the *empirical control.* Sub-section 5.2.1 shows the socio-demographic statistics of the main testing and sub-section 5.2.2 discusses the test score distribution and examines reliability and the dimensionality of the DAC-dimensions. While sub-section 5.2.3 presents the results of the dimensionality and reliability analyses of the DAC-test as a whole, sub-section 5.2.4 describes the findings of the validity analyses. Finally, sub-section 5.2.5 argues the parameters for objectivity.

5.1 CRITERIA FOR VALIDATION

Validity is one of the main goodness criteria for psychometric tests. "[It] is represented in the agreement between two attempts to measure the same trait through maximally different methods" (Campbell & Fiske, 1959, p. 83). As mentioned in sub-section 3.2.4, in modern validity theory *construct validity* dominates other types of validity concepts (e.g. Messick, 1995). For the present research, a correlational approach was chosen to empirically analyse the construct validity of the DAC-test. In this context, Campbell and Fiske (1959) underline examining both sub-types of construct validity: convergent validity and discriminant validity[67]. Therefore, the overall DAC-test score was correlated with the test scores of other theoretically related and unrelated constructs. These constructs and the corresponding selected scales, which build the validation criteria of this research, are presented in sub-sections 5.1.1 - 5.1.6. Each construct is introduced by its definition, followed by the assumed and to-be-tested theoretical hypothesis about the correlation with DAC. Furthermore, the chosen scale, applied for the main testing with its statistical properties, is quoted.

[67] While convergent validity applies to the level to which two measures, which are theoretically supposed to be related, are de facto related, the latter type of validity refers to the degree to which two measures, which are theoretically supposed to be unrelated, are effectively unrelated (e.g. Campbell & Fiske, 1959).

5.1.1 Decision Self-Esteem

Definition of construct

The construct of general *self-esteem* refers to the positive or negative evaluation of the self-concept (e.g. Rosenberg, 1979; Smith & Mackie, 2007), or in other words to the self-assessment of the worth as a person (e.g. MacDonald & Leary, 2012).

"Importantly, self-esteem does not necessarily reflect a person's objective talents and abilities, or even how a person is evaluated by others" (Orth & Robins, 2014, p. 381). The construct involves feelings of confidence and self-acceptance. Regarding *decision self-esteem,* the term alludes to self-confidence and self-reliance in one's competences to successfully solve decision problems.

Theoretical hypothesis about correlation with DAC

"[T]he fact that self-esteem is related to the frequency with which different decision patterns [vigilance, buck-passing, procrastination, and hypervigilance (cf. Mann et al., 1997)] are used (the higher the self-esteem, the less the tendency towards buck-passing and procrastination)" (Sáez de Heredia, Arocena, & Gárate, 2004, p. 116), and findings on the positive correlation of life satisfaction and decision self-esteem (Deniz, 2006) support the hypothesis that a sound style of decision-making referring to DAC positively correlates with decision self-esteem: hence, a significant positive correlation is expected between the final score of the DAC-test and decision self-esteem.

Chosen scale and statistical properties

The Melbourne Decision Making Questionnaire I (henceforth: MDMQ-I; (Burnett, Mann, & Beswick, 1989; Mann, 1982; Mann et al., 1998) is one of the most frequently applied measures for self-esteem in decision science (e.g. Deniz, 2006; Sáez de Heredia et al., 2004) aiming to determine decision-makers' level of self-esteem according to their DMC (e.g. "I think that I am a good decision maker"). The self-rating measure consists of six items, which are graded by assigning numerical values of 1 (not true for me), 2 (sometimes true), or 3 (true for me). Three of the six items are reversed items that are negatively-keyed. After recoding these three items, higher values indicate a higher level of decision self-esteem. Cronbach's alpha of this measure was found to be .74 (Mann et al., 1998). Appendix sub-section 9.11.1 presents the MDMQ-I.

5.1.2 Decision-Making Style

Definition of construct

Scott and Bruce define *decision-making style* as “the learned habitual response pattern exhibited by an individual when confronted with a decision situation. It is not a personality trait, but a habit-based propensity to react in a certain way in a specific decision context” (1995, p. 820). The term is often used in contrast with the term cognitive style (e.g. Thunholm, 2004), which refers to thinking methods applying to decision-making processes (e.g. Hunt et al., 1989). In the framework of the *conflict theory of decision-making* by Janis and Mann (1977), the term decision-making style refers to decision-coping patterns (unconflicted adherence, unconflicted change, defensive avoidance, hypervigilance, and vigilance) that were identified.

Theoretical hypothesis about correlation with DAC

As the vigilant “decision maker clarifies objectives to be achieved by the decision, canvasses an array of alternatives, searching painstakingly for relevant information, assimilates information in an unbiased manner, and evaluates carefully before making a choice ... [and as] vigilance is the only coping pattern that allows sound and rational decision making” (Mann et al., 1997, p. 2), it is assumed for this research that this style of decision-making is related to better DAC-test scores. Conversely, other decision-making styles should be related to lower DAC-test scores.

Chosen scale and statistical properties

The Melbourne Decision Making Questionnaire II (henceforth: MDMQ-II; Mann et al., 1997) was chosen to capture test participants’ decision-making style. The MDMQ-II is the revised version of the *Flinders Decision Making Questionnaire* (Mann, 1982) and consists of 22 items measuring the tendencies to *vigilance* (items 1-6, e.g. “When making decisions I like to collect a lot of information”), *buck-passing* (items 7-12, e.g. “I do not make decisions unless I really have to”), *procrastination* (items 13-17, e.g. “When I have to make a decision I wait a long time before starting to think about it.”), and *hypervigilance* (items 18-22, e.g. “The possibility that some small thing might go wrong causes me to swing abruptly in my preference.”). Like in the MDMQ-I, its response set contains three options; 2-true for me, 1-sometimes true, 0-not true for me. Higher scores refer to a higher accordance of the decision coping pattern. The four sub-scales of MDMQ-II show internal consistency coefficients of α = .80 (vigilance), α = .87 (buck-passing), α = .81 (procrastination), and α = .74 (hypervigilance) (Mann et al., 1998). The items and the instruction for the MDMQ-II are presented in appendix sub-section 9.11.2.

5.1.3 Decision Satisfaction

Definition of construct

Decision satisfaction refers to a cognitive process of judgement whereby decision-makers assess the quality and/or accuracy of their decision after completing the decision-making process. It is an attempt to evaluate the quality of a decision post-decision (Sainfort & Booske, 2000), according to individual decision criteria. In the case of the present study, decision satisfaction describes the assessment of the latest complex decision situation that has been experienced.

Theoretical hypothesis about correlation with DAC

Given the assumption that a sound decision process, i.e. an accurately carried out decision according to decision-analytical guidelines, contributes to a good decision and enhances the chances of achieving a desired outcome (e.g. Hammond et al., 1999; Keren & Bruine de Bruin, 2003; Larrick et al., 1993), it is presumed that persons who reach higher scores in the DAC-test are more satisfied with a recently made decision than persons achieving a lower DAC-test score. Thus, a significant positive correlation of satisfaction with the latest made complex decision and DAC is expected.

Chosen scale and statistical properties

To measure decision satisfaction the Satisfaction with Decision Scale (henceforth: SwD) of Holmes-Rovner et al. (1996) was used. The SwD is based on the conceptual model of an effective decision of O'Connor and O'Brien-Pallas (1989), which, for instance, assumes that the decision-maker is well informed or that the decision is consistent with the decision-maker's objectives. The main aim of the SwD is "to measure global satisfaction with the decision" (Holmes-Rovner et al., 1996, p. 58). The scale consists of six items, e.g. "The decision I made was the best decision possible for me personally", and a five-step response set ranging from strongly disagree, disagree, neither agree nor disagree, agree, to strongly disagree. High scores of the SwD refer to a high satisfaction with one's decisions. Cronbach's alpha of the SwD was found at α =.86 (Holmes-Rovner et al., 1996).

As the SwD was developed to assess patient satisfaction with health care decisions, the instruction had to be slightly adapted relating to the latest complex decision that the person has made. The adjusted instruction and the SwD are presented in sub-section 9.11.3 of the appendix.

5.1.4 Intolerance of Uncertainty

Definition of construct

"Intolerance of uncertainty may be defined as the excessive tendency of an individual to consider it unacceptable that a negative event may occur, however small the probability of its occurrence" (Buhr & Dugas, 2002, p. 932). It describes that people who experience problems in tolerating uncertainty perceive various aspects of life as intolerable due to the presence of risk, uncertainty, and ambiguity in daily situations. "Intolerance of uncertainty can be seen as a filter through which individuals view their environment", explain Buhr and Dugas (2002, p. 933).

Theoretical hypothesis about correlation with DAC

Dugas and Robichaud (2012) describe three thinking processes that might appear when individuals, intolerant of uncertainty, are facing uncertainty in decision situations. In comparison to individuals who tolerate uncertainty, these people: (1) perceive the ambiguous situation as more threatening, (2) need more information before making a decision, and (3) feel less confident making a decision. "It could, thus, be argued that [intolerance of] uncertainty ... represents procrastination of action until sufficient information is obtained, or of decision-making due to a lack of confidence" (Birrell, Meares, Wilkinson, & Freeston, 2011, p. 1205). As complex decisions are described as uncertain and as the DAC-test aims to capture UNCERTAINTY, it could be assumed that people who do not tolerate uncertainty (e.g. J. E. Matheson & Howard, 1968), perform worse in sound decision-making assessed by the DAC-test. Thus, for this Ph.D. research a negative correlation is expected between the measure of intolerance of uncertainty and the DAC-test score.

Chosen scale and statistical properties

To capture test participants' intolerance of uncertainty, the Intolerance of Uncertainty Scale (henceforth: IoU) of Buhr and Dugas (2002) was selected. "The original French version of the IoU was developed to assess emotional, cognitive, and behavioral reactions to ambiguous situations, implications of being uncertain, and attempts to control the future" (Buhr & Dugas, 2002, p. 934). The corresponding English measure of Buhr and Dugas (2002) consists of 27 items with a five-point Likert scale ranging from 1 (not at all characteristic of me) to 5 (entirely characteristic of me). A promax-rotated principal-factor analysis had identified four moderately correlated factors (λ [.42; .69], $p_s < .001$), which capture: *uncertainty leads to the inability to act, uncertainty is stressful and upsetting, unexpected events are negative and should be avoided, being uncertain about the future is unfair*. The correlation with the IoU overall score varies from $r_P = .82$ to $r_P = .94$ ($p < .001$). For the present study, the sub-scale *uncertainty leads to the inability to act* was chosen to measure the decision relevant aspect of intolerance of uncertainty, e.g. "When I am uncertain, I can't go forward". Its Cronbach's alpha was found to be $\alpha = .87$. The eight items of the sub-scale and the scale instruction are presented in sub-section 9.11.4 of the appendix. High scores of the IoU sub-scale refer to a high degree of IoU.

5.1.5 Fluid Intelligence

Definition of construct

Fluid intelligence is a factor of *general intelligence*, which was established by Cattell (1963, 1987) and Horn (1985). It describes the ability to think logically and solve problems. It captures inductive and deductive thinking, independent of the acquired knowledge. The construct is correlated with success factors in areas like education (e.g. Kuncel, Hezlett, & Ones, 2004; Kunina, Wilhelm, Formazin, Jonkmann, & Schroeders, 2007) and career (e.g. Hunter & Hunter, 1984; Ng, Eby, Sorensen, & Feldman, 2005; Salgado et al., 2003).

Theoretical hypothesis about correlation with DAC

As analytical decision-making captured by the DAC-test is defined as a quantitative and logical procedure to solve decision problems - a procedure that allows the "logical balancing of the factors that influence a decision" (Howard, 1973, p. 64), a positive correlation of a measure for fluid intelligence and the DAC-test score is expected here. Comparably, Bruine de Bruin and colleagues (2007) have found a significant positive correlation of their A-DMC test and Raven's standard progressive matrices (Raven, Raven, & Court, 2003) of $r_P = .61$ (p < .001).

Chosen scale and statistical properties

Raven's standard progressive matrices are often applied to measure fluid intelligence (e.g. Fry & Hale, 1996; Hayashi, Kato, Igarashi, & Kashima, 2008). However, for the present study a less time consuming, online, and open access scale had to be found as the overall length of the testing procedure, including the DAC-test and its validation criteria, was intended to be kept as short as possible to reduce test participants' workload. Thus, the short version of the Hagen Matrices Test (henceforth: HMT-S;

Heydasch, Haubrich, & Renner, 2013) was chosen. It intends to measure *fluid reasoning* and more specifically, *induction* as the core aspect of fluid intelligence (cf. Schneider & McGrew, 2012). The HMT-S consists of six matrices with an internal consistency of α = .64 and a correlation of r_P = .49 (p < .001) with the *Intelligenz-Struktur-Test 2000R* (Liepmann, Beauducel, Brocke, & Amthauer, 2007; [engl: inteligence-structure-test]). In the present study the Cronbach's alpha was found at α = .70. On average it takes nine minutes to complete the HMT-S. Performance is measured by the number of matrices answered correctly. Appendix sub-section 9.11.5 presents the instruction and an example matrix of the HMT-S.

5.1.6 Problem-Solving Competence

Definition of construct

"Problem solving is a process in which we perceive and resolve a gap between a present situation and a desired goal. ... In general, the situation is one not previously encountered, or where at least a specific solution from past experiences is not known." (Huitt, 1992, p. 34) In comparison to *reasoning, problem-solving competence* focuses on the "experimental interactions with the environment to clarify the nature of a problem and potential solutions" (Raven, 2000, p. 54). Therefore, the task environment is dynamic and some of the corresponding "regularities can only be revealed by successful exploration and integration of the information gained in that process" (Axel Buchner in Frensch & Funke, 1995, p. 14). In this framework, problem-solving competence describes the ability to observe and examine the variables of a situation, in which the status quo and the desired goal differ, and to develop a strategy to move towards achieving the defined goal.

Theoretical hypothesis about correlation with DAC

"The steps in both problem solving and decision making are quite similar. In fact, the terms are sometimes used interchangeably" (Huitt, 1992, p. 34). In the present research, the term DAC refers to the cognitive and motivational steps of the decision-making process and omits its voluntary phase (cf. sub-section 2.4.3). Thus, DAC measured by its corresponding test is seen as one part of the general problem-solving process, which starts with an input phase and ends with the implementation and reviewing phase (e.g. Bransford & Stein, 1984). Inferentially, a positive correlation is expected between participants' DAC-test score with the measure for problem-solving competence for this research.

Chosen scale and statistical properties

To capture problem-solving competence an online version of the Tower of Hanoi (henceforth: ToH) has been selected. The puzzle developed by the French mathematician Èdouard Lucas in 1883 (Hinz, 1992) has been used in experimental settings for decades (e.g. Anderson & Douglass, 2001; H. J. Bell, 2004; Schiff & Vakil, 2015; Vernon & Strudensky, 1988).

For the present study the chosen ToH version consists of four rings with different sizes and three rods. All rings are placed on the rod in the middle. The main aim of the task is to move all rings to one of the other rods in as few moves as possible. At the end the rings have to be sorted by size; starting with the largest ring at the bottom. Performance is measured by the number of total moves – the fewer moves the better. The task instruction and an exemplification are presented in sub-section 9.11.6 of the appendix.

5.1.7 Socio-Demographic Variables

Additional to the validation criteria presented in sub-sections 5.1.1 to 5.1.6, age, grade point average, and average mark in maths as socio-demographic variables have been surveyed.

Age

Previous studies on age-related differences of behavioural DMC reveal mixed results (cf. Bruine de Bruin et al., 2012; Hanoch, Wood, & Rice, 2007; Peters, Hess, Västfjäll, & Auman, 2007): Some decision-making abilities seem to increase with age, such as resisting the influence of irrelevant alternatives (e.g. Kim & Hasher, 2005; Tentori, Osherson, Hasher, & May, 2001). Other skills seem to decrease with age, such as the application of decision rules (cf. Bruine de Bruin et al., 2007). And yet other tasks, such as resisting to framing effects, show ambiguous results in various studies (cf. Finucane, Mertz, Slovic, & Schmidt, 2005 v. Rönnlund, Karlsson, Laggnäs, Larsson, & Lindström, 2005).

In general, behavioural decision science scholars conjecture that the decision-making performance of older adults depends on a decision-making task's underlying required cognitive ability. While fluid cognitive skills begin to weaken in the early 20s (e.g. Park et al., 2002; Salthouse, 2004), crystallized cognitive abilities profit from life experiences that come with age (e.g. Neubauer, 2005). Also, emotion-related abilities, such as emotion regulation (e.g. Mather & Carstensen, 2003) and recognising emotional states (e.g.

Labouvie-Vief, DeVoe, & Bulka, 1989), show evolvement over the lifespan (e.g. Forstmeier, Uhlendorff, & Maercker, 2005).

So far, previous studies have mostly compared age groups that reflect different stages of psychological development (cf. Erikson, 1959); ranging from adolescence or early adulthood to maturity (e.g. Bruine de Bruin et al., 2007, 2012). The present study in contrast focuses on the age group of young adults between 18 and 30 years and thus explores the potential age-related variations within one phase of psychological development.

Given the mixed study results of age-related performances on decision-making tasks, the variety of required decision tasks presented by the DAC-test (cf. section 4.1), and the relatively tight-scoped age range of the sample group of the present study (cf. sub-section 3.3.1, it is presumed that age and the DAC-test performance do not significantly correlate.

Grade point average

Several studies have shown correlations between DMC and various measures of intelligence-related aspects as e.g. fluid intelligence and numeracy (cf. Bruine de Bruin et al., 2007; Del Missier, Mäntylä, & Bruine de Bruin, 2010; Dieckmann, 2008), crystalline intelligence (Bruine de Bruin et al., 2007; Finucane & Gullion, 2010), or specialised knowledge (Stanovich, 2009; cf. Stanovich & West, 2008). As intelligence in general is considered a key predictor for academic success (cf. Neisser et al., 1996) within this study a positive correlation between the grade point average of study participants and the DAC-test score is assumed.

Average mark in maths

Numeracy, which describes “the ability to understand and work with numbers” (Oxford Dictionary Online, n.d., [entry “numeracy”]) has been identified as an influencing factor to decision making (cf. Peters et al., 2006; Reyna, Nelson, Han, & Dieckmann, 2009). Thus, it is expected that the average mark in maths correlates positively with DAC in general, but especially with those DAC-dimensions that rely on numeracy such as UNCERTAINTY (cf. sub-section 4.1.6).

5.1.8 Pre-tested & Rejected Criteria for Validation

Besides the above listed validation criteria, other potentially relevant constructs have been considered and corresponding scales have been surveyed and analysed in the pre-testing. However, the Generalised Self-Efficacy Scale of Schwarzer and Jerusalem (1995), the Satisfaction with Life Scale of Diener, Emmons, Larsen, and Griffin (1985), and the Decision Outcome Inventory (Bruine de Bruin et al., 2007) were tested and rejected as having no significant correlation; neither with the overall DAC-test score nor with any of the DAC-dimensions. Also, the rejection served to reduce study participants' workload for the main testing.

The following section presents the results of the main testing and thus, shows how the finally selected validation criteria perform.

5.2 RESULTS OF THE EMPIRICAL CONTROL

The DAC-test, which was adapted after the pre-testing (cf. section 4.2) and is presented in appendix section 9.8, and the criteria for validation introduced in section 5.1 were surveyed in the main testing phase of the present Ph.D. research

The results of the so-called *empirical control*, which serves to analyse the goodness criteria of the developed DAC-test, are presented in the following sub-sections. While sub-section 5.2.1 gives an overview of the socio-demographics of this main study, sub-section 5.2.2 discusses the test score distributions and examines the dimensionality of the sub-scales of the DAC-test, and in sub-section 5.2.3 the dimensionality of the construct and its internal consistency is analysed. Sub-section 5.2.4 reveals the results of the test's validation and sub-section 5.2.5 reviews the arguments for the test's objectivity.

5.2.1 Socio-Demographics of the Main Testing

In 2015, the sample group for the main testing (cf. corresponding to the criteria for the target group in sub-section 3.3.1) was acquired through five institutions: volunteers of the acquisition list for study participants of the Max Planck Institute of Human Development in Berlin, a group of students taking the "Judgement and Decision Making for Management" summer school course at the London School of Economics and Political Science, students of "Industrial and Systems Engineering" course at Virginia Tech, voluntary students of the social and psychological department at Free University of Berlin, and voluntary students of the management department of University of Bayreuth.

Overall, 455 people did participate in the main study of the present research. 368 of these were selected for the statistical analyses. According to the pre-testing, three selection criteria caused this reduction to 80.9% of the total sample: (1) participants younger than 18 or older than 30 years, (2) participants with English skills that were self-rated as “basic English knowledge”, and (3) test processing times shorter than 20min[68] and longer than 3h.

The percentage distribution of study participants by their acquisition source is presented in Table 5-1. Twenty eight (7.6%) test participants were sent from the University of Bayreuth, 37 (10.1%) came from Free University of Berlin, 63 (17.1%) students came Virginia Tech, 66 (17.9%) students were from the London School of Economics and Political Science, 80 (21.7%) were contacted via the mailing list of the Max Planck Institute for Human Development, and 94 (25.5%) persons came from multiple other sources, such as former students of the Free University of Berlin and Humboldt University of Berlin, who had heard about the test or colleagues and friends of the research group.

Table 5-1 Distribution by source of acquisition of the main testing sample

Source	Frequency	%
UB	28	7.6
FU	37	10.1
VT	63	17.1
LSE	66	17.9
MPI	80	21.7
Misc.	94	25.5
Total	368	100.0

Note. UB = University of Bayreuth; FU = Free University of Berlin; VT = Virginia Tech; LSE = London School of Economics and Political Science; MPI = Max Planck Institute for Human Development Berlin, Misc. = miscellaneous.

The mean age of the sample group was 24 years (M = 23.75; SD = 3.41; Range = 18-30). About 49.7% of the 364 test participants[69] assessed their English skills as “mother tongue” or “very good/business fluent” (cf. Table 5-2). The remaining 50.3% saw their language skills as “good/conversant” or “fluent”.

[68] In contrast to the pre-testing, data sets with processing times between 20min and 30min were kept for the statistical analyses as they showed proper results in terms of DAC-test performance – results better than the probability of guessing the answers correctly. The achieved performance in the DAC-test is treated as an indicator for thoroughness of participation.

[69] Data contains four missing data sets.

Table 5-2 Frequencies of self-assessed English skills of the main testing sample

English level	Frequency	%	Valid %
Good/conversant	55	14.9	15.1
Fluent	128	34.8	35.2
Very good/business fluent	113	30.7	31.0
Mother tongue	68	18.5	18.7
NA	4	1.1	
Total	368	100.0	

Note. NA = missing data/not available.

Three hundred and sixty six individuals of the study participants (99.5%) specified their gender: One hundred and seventy four (47.5%) of the sample group were male and 192 (52.5%) were female.

Of the main testing sample group, 152 participants (41.3%) received a financial incentive of €15.00. The remaining 216 participants (58.7%) were not paid and participated either voluntarily or because the DAC-test was presented as part of an academic course they took at that time.

The greatest national group of the main study were Germans with 52.9%, followed by Americans with 16.4%, Indians with 7.4%, Chinese with 3.8%, Polish with 1.6%, and British and South-Korean each with 1.4% of the valid data (365 test participants, cf. Table 5-3).

Table 5-3 Distribution by nation of the main testing sample

Nationality	Frequency	%	Valid %
German	193	52.4	52.9
American	60	16.3	16.4
Indian	27	7.3	7.4
Chinese	14	3.8	3.8
Polish	6	1.6	1.6
British	5	1.4	1.4
South-Korean	5	1.4	1.4
Misc.	55	14.9	15.0
NA	3	0.8	
Total	368	100.0	

Note. Misc. = miscellaneous; NA = missing data/not available; National groups consisting of less than three individuals have been summarised under misc.

The average processing time of the main testing for taking the DAC-test and responding to the criteria for validation averaged out at 97 minutes (cf. Table 5-4). Completing only the case study test, participants took around 73 minutes.

Table 5-4 Processing times (main testing)

Processing time...	*N*	*Min*	*Max*	*M*	*SD*	*Mdn*	*IQR*
...for the case	368	16	162	73.12	29.85	65.82	31.44 - 81.47
...for the case and criteria of validity	368	22	178	97.33	33.17	85.94	39.43 - 128.17

How the sample group of the main testing stated their highest educational level is presented by Table 5-5: Out of the 358 test participants who specified this information, 1.4% indicated not holding any school leaving certificate. Thirteen point four per cent gained a certificate after nine or 10 years of school education, 33.5% hold a university-entrance diploma, 35.2 have completed their Bachelor studies, 14.8% their Diploma or Master studies, and 1.7% hold a doctoral degree.

Table 5-5 Distribution by degree of education of the main testing sample

Educational certificate	Frequency	%	Valid %
No school leaving certificate	5	1.4	1.4
Certificate after nine or 10 years of school education	48	13.0	13.4
University-entrance diploma	120	32.6	33.5
Bachelor's degree	126	34.2	35.2
Diploma or Master's degree	53	14.4	14.8
Ph.D.	6	1.6	1.7
NA	10	2.7	
Total	368	100.0	

Note. NA = missing data/not available.

Table 5-6 shows how test participants of the main testing have answered the item "What is your former, current or intended profession? Which field does it belong to?": The four largest groups of the 323 test participants who stated this information come from Business and Financial Operations Occupations with 21.1%, from Architecture and Engineering Occupations with 16.4%, from Life, Physical, and Social Science Occupations with 10.5%, and from Healthcare Support Occupations with 10.2%.

Table 5-6 Distribution by former, current or intended profession of the main testing sample

Profession	Frequency	%	Valid %
Architecture and Engineering Occupations	53	14.4	16.4
Arts, Design, Entertainment, Sports, and Media Occupations	24	6.5	7.4
Building and Grounds Cleaning and Maintenance Occupations	1	0.3	0.3
Business and Financial Operations Occupations	68	18.5	21.1
Community and Social Services Occupations	7	1.9	2.2
Computer and Mathematical Occupations	21	5.7	6.5
Construction and Extraction Occupations	2	0.5	0.6
Education, Training, and Library Occupations	26	7.1	8.0
Food Preparation and Serving Related Occupations	3	0.8	0.9
Healthcare Support Occupations	33	9.0	10.2
Installation, Maintenance, and Repair Occupations	1	0.3	0.3
Legal Occupations	5	1.4	1.5
Life, Physical, and Social Science Occupations	34	9.2	10.5
Management Occupations	27	7.3	8.4
Office and Administrative Support Occupations	6	1.6	1.9
Personal Care and Service Occupations	3	0.8	0.9
Production Occupations	4	1.1	1.2
Sales and Related Occupations	2	0.5	0.6
Transportation and Material Moving Occupations	3	0.8	0.9
NA	45	12.2	
Total	368	100.0	

Note. NA = missing data/not available.

5.2.2 Test Score Distribution, Reliability, & Dimensionality of the DAC-Dimensions

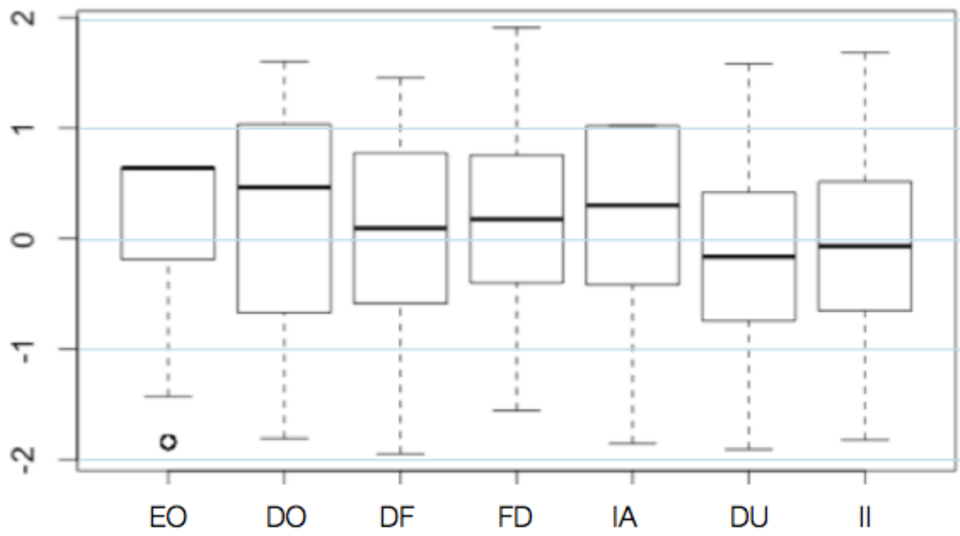
A basic condition for running specific statistical analyses with interval and ratio scaled data is the normal distribution of the test data set. To investigate for normality, the values of skewness and kurtosis (cf. Table 5-7) and graphical distributions (cf. Figure 5-1 and Table 5-8) were examined. Table 5-7 presents the descriptive statistics for the calculated sum scores of the seven operationalized DAC-dimensions. Comparable to the results of the pre-testing, all DAC-test sub-scales besides OBJECTIVES conform to the acceptable range between -1 and +1 of skewness and kurtosis (George & Mallery, 2011) for proving normal univariate distribution.

Table 5-7 Descriptive statistics of the sum scores of the seven operationalized DAC-dimensions

DAC-dimension	*M*	*SD*	Potential range	Observed range	Skew	Kurtosis	*α*
EO sum score	8.45	2.42	0-10	0-10	-1.68	2.14	-
DO sum score	5.18	1.76	0-8	0-8	-0.50	-0.46	.58
DF sum score	4.86	1.47	0-7	0-7	-0.38	-0.27	.47
FD sum score	3.70	1.73	0-8	0-8	0.01	-0.53	.42
IA sum score	5.58	1.39	0-7	1-7	-0.88	0.16	.51
DU sum score	3.28	1.72	0-7	0-7	0.09	-0.73	.61
II sum score	6.24	3.42	0-12	0-12	0.21	-0.88	.86

Note. *N* = 368; DAC = Decision-analytical competence; skew = skewness, *α* = Cronbach's alpha; EO = Ability to envision one's objectives; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

The scaled boxplots in Figure 4-14 (*M* = 0; *SD* = 1; statistical outliers of two or more standard deviations are excluded) present a comparable view of the test score distributions: Complementary to the skewness values of OBJECTIVES the OBJECTIVES-boxplot reveals that the median of the OBJECTIVES values closes the upper quartile and thus, builds the upper extreme of this test score distribution. This means that more than half – more precisely 56.79% – of the test participants achieved the highest score of this DAC-dimension; the scores plateau.



Note. EO = Ability to envision one's objectives; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information

Figure 5-1 Boxplots of the seven scaled DAC-dimensions (main testing)

As happened in the pre-testing, a ceiling effect (cf. Zedec, 2014) appeared. This effect does not allow for measuring variance within the sample group. For this reason, OBJECTIVES had to be excluded from the analyses of the main testing.

Perusing the main testing data of OBJECTIVES, test participants generated 10 objectives (*M* = 9.73; *SD* = 6.57; range: 0 - 30) on average. They identified a mean of 23 objectives as relevant from the master list (*M* = 22.87; *SD* = 7.72; range: 1 - 40) and matched 46.55%[70] of the recognised objectives on average (*SD* = 24.09%; range: 0% - 100%) with their self-generated objectives. The difference of 13.14 objectives between 9.73 self-generated objectives and 22.87 recognised objectives leads to the assumption, similar to the findings of Bond and colleagues (2008) and to the results of the pre-testing, that test participants either had problems to envision a complete list of individual relevant objectives or that they did not put enough effort into completing the first task, or a combination of both.

In comparison to the pre-testing, participants of the main testing created 1.78 objectives more in the first step of the task. Also, on average 2.99 more objectives of the master list were recognised in the main testing and the percentage of matched objectives increased from 31.15% to 45.55%. It could also be presumed that test participants of the main testing in comparison to participants of the pre-testing were either more skilled, more motivated, and/or more concentrated when completing the matching task of OBJECTIVES.

While participants of the pre-testing envisioned on average 78.57% of their seven chosen most important objectives, in the main testing participants selected on average 84.50% of their self-generated objectives for the list of the 10 most important objectives. In percentage terms, this means more self-generated objectives in the main testing, independent of the change to the task instruction (cf. sub-section 4.2.5).

In all four steps of the OBJECTIVES task (cf. sub-section 4.1.1) participants of the main testing have performed better than participants of the pre-testing. The increase from seven (pre-testing) to 10 (main testing) required objectives in the last step of the OBJECTIVES task did not help in terms of the observed ceiling effect. The OBJECTIVES task per se seems to be too easy for test participants. Potential explanations and ideas for improvement are discussed in section 6.1.

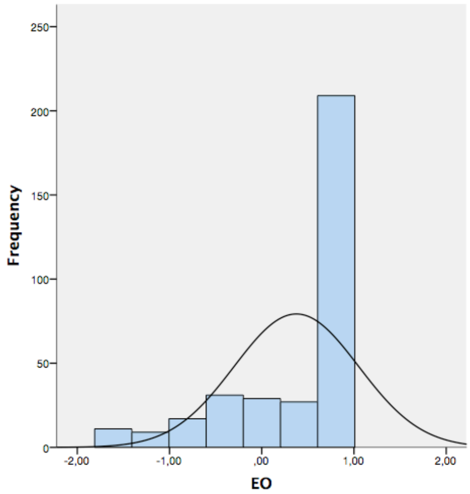
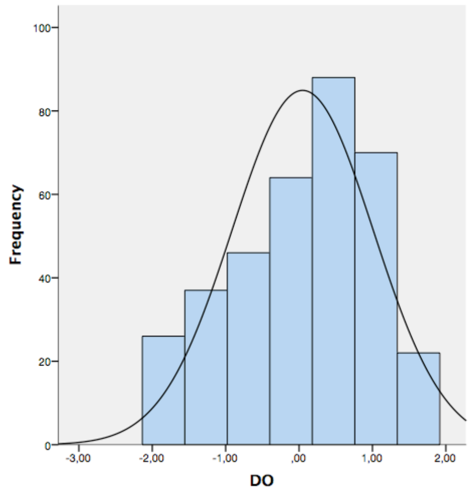
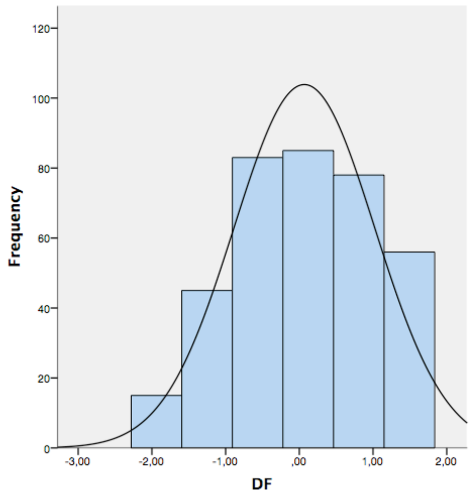
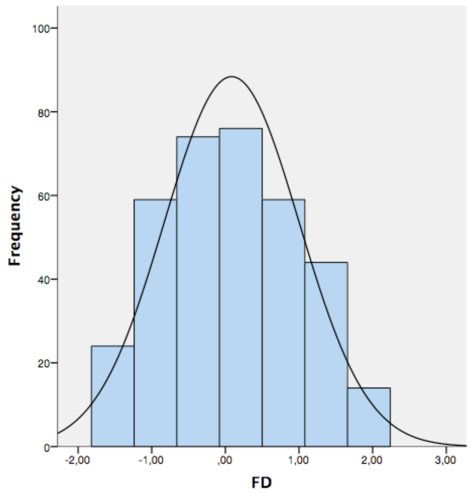
Once again, the discrepancy of OBJECTIVES' distribution from normal distribution can be observed by its sum score distribution with overlaid curve of a normal distribution (cf. very high row, left column of Table 5-8). The values of OBJECTIVES present an extreme in

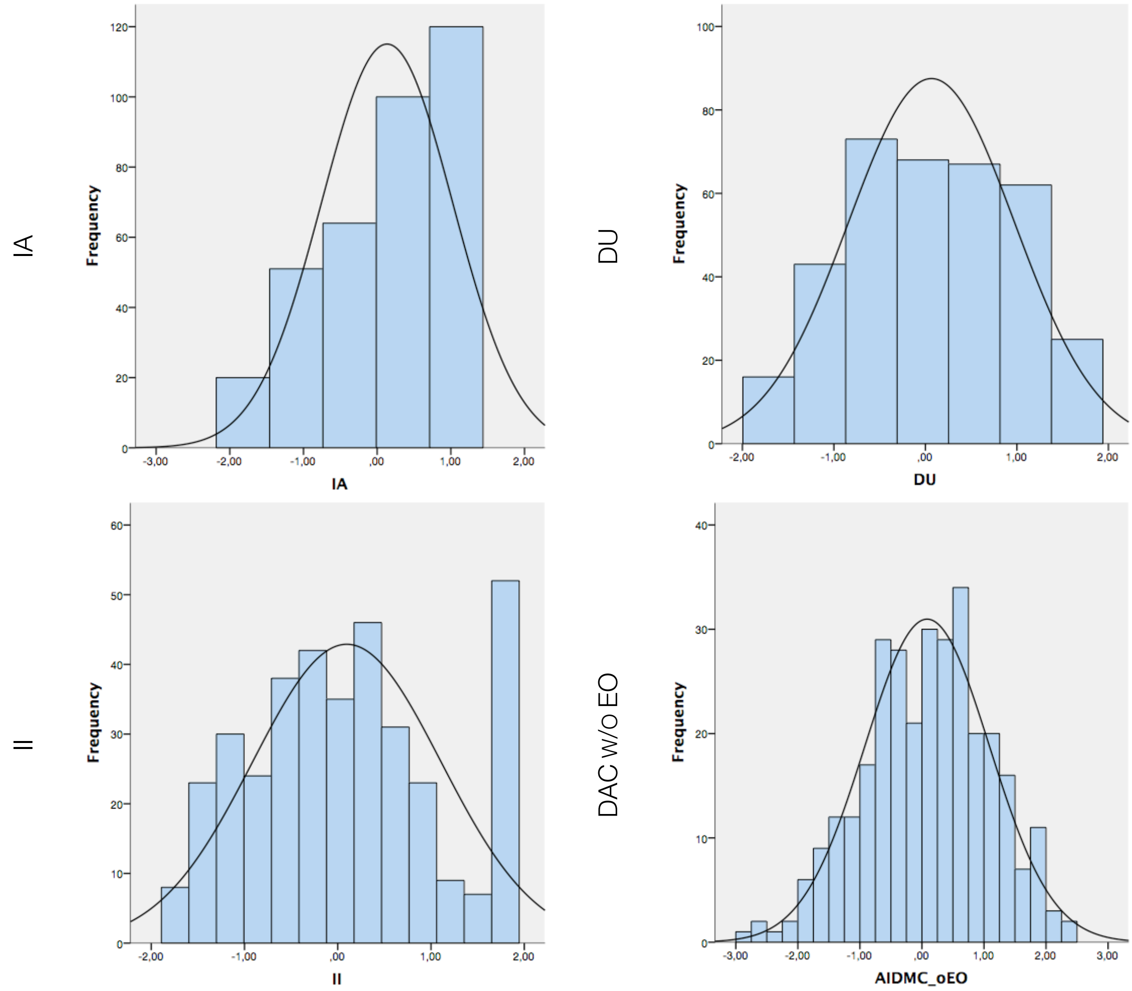
[70] Due to the matching of self-generated objectives and objectives presented at the master list, it is possible to match one self-generated objective to several objectives from the master list, and also the other way around.

the area of high scores. Table 5-8 also presents the histograms for the other six DAC-dimensions, each with an overlaid curve of a normal distribution.

As the values of OBJECTIVES violate the condition of a normal distribution they had to be excluded from further analyses. Therefore, the left cell of the very low row of Table 5-8 presents DAC-test sum score distribution of the main testing without the OBJECTIVES' values. According to the display of this frequency distribution, the assumption of a normal distribution of the main testing data set is justified.

Table 5-8 Histogram of each DAC-dimension and the DAC-overall-score (main testing)

	Histogram of scaled values		Histogram of scaled values
EO		DO	
DF		FD	



Note. DAC = Decision-analytical competence; EO = Ability to envision one's objectives (*N* = 333; *M* = 0.38; *SD* = 0.68); DO= Ability to recognise decision opportunities (*N* = 353; *M* = 0.49; *SD* = 0.96); DF = Ability to assess decision fitness (*N* = 362; *M* = 0.07; *SD* = 0.95); FD = Ability to frame a decision (*N* = 350; *M* = 0.09; *SD* = 0.92); IA = Ability to identify relevant alternatives (*N* = 355; *M* = 0.13; *SD* = 0.89); DU = Ability to deal with uncertainty (*N* = 354; *M* = -0.07; *SD* = 0.91); II = Ability to integrate information (*N* = 368; *M* = 0.1; *SD* = 1.01); DAC w/o EO = Decision-Analytical Competence without ability to envision one's objectives (*N* = 312; *M* = 0.8; *SD* = 1.01).

[a] Statistical outliers were excluded.

To examine the internal consistency for each DAC-test sub-scale, Cronbach's alpha and McDonald's omega (McDonald, 1999)[71] were calculated. Table 5-9 compares the reliability coefficients for each DAC-test sub-scale of the pre-testing to the main testing and specifies the number of items that have been considered for this analysis.

[71] McDonald's omega (ω_H) is an estimate for internal consistency and captures the general factor saturation of a scale. In comparison to Cronbach's alpha, which purely relies on the common variance of all items, McDonald's omega is based on factor loadings and thus presents a different picture of a test's internal consistency (McDonald, 1999). The desired threshold for McDonald's omega is $\omega_H \geq .70$ (Zinbarg et al., 2005). As Cronbach's alpha assumes tau equivalent items (Cronbach, 1951), it only suits scales where all items have comparable discriminatory power (cf. the range of discriminatory power coefficients in the main testing in appendix section 9.12); the more sophisticated measure called McDonald's Omega was used.

Table 5-9 Internal consistencies of the DAC-dimensions (pre-testing & main testing)

DAC-dimension	Pre-testing (pre-testing)		Main testing (main testing)		
	α	Number of items *after* item selection	α	ω_H	Number of items *after* item selection
DO sum score	.63	9	.58	.76	8
DF sum score	.33	8	.47	.66	7
FD sum score	.51	8	.42	.56	8
IA sum score	.44	8	.51	.69	7
DU sum score	.64	8	.61	.84	7
II sum score	.85	12	.86	.97	12

Note. N = 143 (pre-testing) & 368 (main testing); DAC = Decision-analytical competence; α = Cronbach's alpha; ω_H = McDonald's omega; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

To examine the dimensionality of each DAC-dimension, CFAs were run to test for homogeneity. Therefore, one-factor solutions for each sub-scale were tested separately. The analyses were carried out in R with the lavaan-toolbox (http://lavaan.ugent.be) using a maximum likelihood estimator with "Huber-White" robust standard errors and a (scaled) robust test statistic (Rosseel, 2012).

As the Chi-square (χ^2) statistic tends to be susceptible to sample size, i.e. χ^2 was found to become significant for even small differences in large sample groups (cf. Hair et al., 1998; Hoyle, 1995; Netemeyer, Bearden, & Sharma, 2003), other model fit indices as the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Weighted Root-Mean-Square Residual (WRMR) were additionally consulted. For evaluating the DAC sub-scales for homogeneity the cut-off criteria of Brown and Cudeck (1993), Hu and Bentler (1999), and Yu (2002) were used; CFI ≥ .95, RMSEA ≤ .08, WRMR ≤ 1. Table 5-10 gives an overview of the various fit indices for the homogeneity tests.

Table 5-10 CFAs to test DAC-test sub-scales for homogeneity

DAC-dimension	χ^2	*df*	WRMR	CFI	RMSEA	CIRMSEA
DO sum score	81.341***	20	1.362	.975	.092	.072 - .114
DF sum score	12.330	14	0.577	1.000	.000	.000 - .045
FD sum score	27.392	20	0.804	.924	.032	.000 - .059
IA sum score	18.893	14	0.735	.975	.031	.000 - .063
DU sum score	89.993***	14	1.708	.893	.122	.099 - .147
II sum score	706.295***	54	3.733	.987	.181	.170 - .193

Note. N = 368; CFA = Confirmatory factor analysis; DAC = Decision-analytical competence; χ^2 = Chi-square; *df* = degrees of freedom, WRMW = Weighted Root-Mean-Square Residual; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CIRMSEA = Confidence interval of RMSEA; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Of the nine items of OPPORTUNITIES, one item (item DO_01) had to be deleted as in the split-half analysis[72] it appeared to be unintentionally negatively poled[73]. Cronbach's alpha for the remaining eight items is .58, and therefore does not meet the desired threshold for this quality criterion. As discussed in sub-section 4.2.5, a low value for Cronbach's alpha can be seen as an indicator for multi-dimensionality. Correspondingly, a confirmatory factor analysis for binary data was run to examine the scale for homogeneity by a one-factor model. Table 5-9 gives an overview of the weighted root-mean-square residual (WRMR, cf. Yu, 2002) for each DAC-dimension. A value above 1 for the WRMR reveals that the scale for OPPORTUNITIES is not homogeneous, χ^2 (20, N = 368) = 81.341, $p < .001$, WRMR = 1.362, CFI = .975, RMSEA = .092, 90%-CIRMSEA[74] = .072 - .114. To further analyse internal consistency of the sub-scale for OPPORTUNITIES, McDonald's omega was calculated and met the desired threshold with $\omega_H = .76$.

FITNESS was measured with an eight-item scale in the main testing. Due to one negatively loading item (item DF_02) in the split-half analysis, seven items remained for further analyses. The internal consistency of the seven-item measure shows a Cronbach's alpha of .47, which is below the aspired critical value of .70. Here, the confirmatory factor analysis supported the assumption for the scale of FITNESS being homogeneous, χ^2 (14, N

[72] Split-half reliability is "a measure of the internal consistency of surveys, psychological tests ... [and] is determined by dividing the total set of items (e.g., questions) relating to a construct of interest into halves (e.g., odd-numbered and even-numbered questions) and comparing the results obtained from the two subsets of items thus created." (Zedec, 2014, p. 346 [entry "split-half reliability"])

[73] Decisive for the deletion of this item after the final item selection was that no item in the pre-testing appeared to be negatively-poled. The unexpected negative-poling in the main testing led to the unusual exclusion of this item and three other items; one from each dimension: FITNESS, ALTERNATIVES, and UNCERTAINTY.

[74] CIRMSEA = Confidence interval of RMSEA

= 368) = 12.330, p = .580, WRMR = 0.577, CFI = 1.000, RMSEA = .000, 90%-CIRMSEA = .000 - .045. McDonald's omega for the DAC sub-scale of FITNESS slightly missed the corresponding threshold with ω_H = .66.

Eight items measured the DAC-dimension FRAME. The Cronbach's alpha for this scale is .42 and is below the critical threshold. A confirmatory factor analysis for binary data confirmed that the items of FRAME form a homogeneous scale, χ^2 (20, N = 368) = 27.392, p = .125, WRMR = 0.804, CFI = .924, RMSEA = .032, 90%-CIRMSEA = .000 - .059. With ω_H = .56 the critical value for the internal consistency of FRAME analysed by McDonald's omega was missed.

One (item IA_05) of the eight items of ALTERNATIVES had to be removed from further analyses, as it appeared to be unintentionally negatively poled in the split-half analysis. The remaining seven items show an internal consistency of α = .51, which is below the critical value for Cronbach's alpha. The remaining seven items build a homogeneous scale for ALTERNATIVES, χ^2 (14, N = 368) = 18.893, p = .169, WRMR = 0.735, CFI = .975, RMSEA = .031, 90%-CIRMSEA = .000 - .063. For the DAC sub-scale of ALTERNATIVES, McDonald's omega also met the threshold by ω_H = .69.

As for OPPORTUNITIES, FITNESS, and ALTERNATIVES, one item (item DU_01) of UNCERTAINTY had to be removed as it was negatively poled in the split-half analysis. The remaining seven items of UNCERTAINTY show a Cronbach's alpha of .61. The WRMR above the threshold of 1 disproves the assumption of homogeneity for the scale of UNCERTAINTY, χ^2 (14, N = 368) = 89.993, p < .001, WRMR = 1.708, CFI = .893, RMSEA = .122, 90%-CIRMSEA = .099 - .147. In terms of McDonald's omega, the threshold of internal consistency was met by ω_H = .84.

Also for INFORMATION Cronbach's alpha was applied to measure internal consistency. The 12 items of this DAC-dimension present a reliability coefficient of .86, which meets the desired threshold for Cronbach's alpha. The 12 items of INFORMATION generate a heterogeneous scale, χ^2 (54, N = 368) = 706.295, p < .001, WRMR = 3.733, CFI = .987, RMSEA = .181, 90%-CIRMSEA = .170 - .193. McDonald's omega turned out to be very high at ω_H = .97. Looking at the results of validity and reliability analyses on the DAC-dimension level, it is conspicuous that the three heterogeneous sub-scales present McDonald's omegas over .70, while the homogeneous sub-scales miss this threshold. The apparent difference between the evaluations of validity examined by the CFAs and reliability measured by internal consistency might have resulted from the fact that, on the one hand CFA tests are only unidimensional and discriminatory power of the items do not influence model fit. Measures of reliability on the other hand refer (in part) to those very discriminatory powers. Thus, results can be diverging, just as validity and reliability measure different aspects of a test. In the case of e.g. FITNESS, the inter-item correlations are relatively low, even though the CFA confirms its homogeneity. So, the items only roughly measure the same construct. Due to the low inter-item correlations a second factor is not formed.

5.2.3 Reliability & Dimensionality of the DAC-Test

To further research the dimensionality of the DAC-dimensions, Table 5-11 shows the bivariate correlations among the six remaining DAC-test sub-scales. All correlations are significant and positive, and range from low r_p = .12 (among UNCERTAINTY and OPPORTUNITIES) to moderate r_p = .44 (among INFORMATION and UNCERTAINTY) with an overall mean correlation of r_p = .24.

Table 5-11 Pearson correlations (r_p) between DAC-test sub-scales

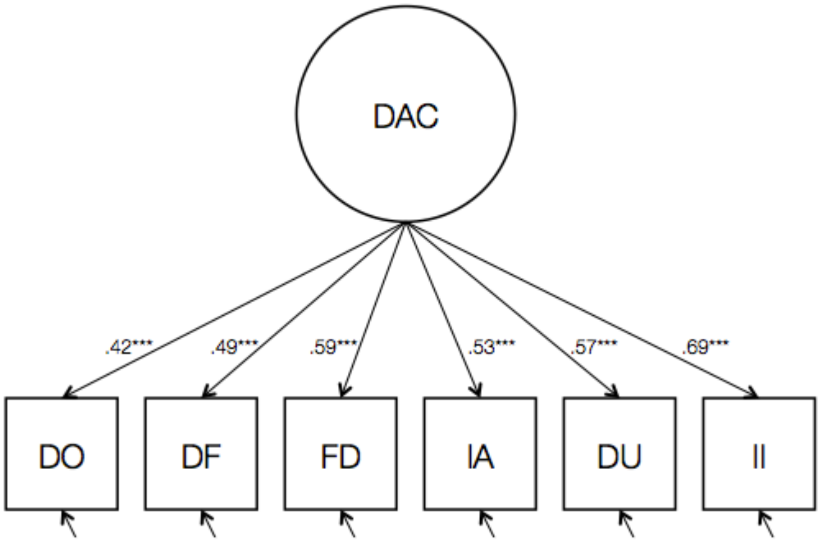
	DO	DF	FD	IA	DU	II
DO	1					
DF	.27***	1				
FD	.27***	.20***	1			
IA	.19***	.23***	.23***	1		
DU	.12*	.16**	.31***	.22***	1	
II	.17**	.23***	.33***	.34***	.44***	1

Note. DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

[a] Mean correlation: r_p = .24.

*Two-sided; *$p < .05$; **$p < .01$; ***$p < .001$.

Based on decision-analytical theory (cf. chapter 2), the corresponding research hypotheses (cf. section 3.1), and the results of the EFA in the pre-testing (cf. sub-section 4.2.4), a CFA (cf. Figure 5-2) was run to confirm the one factor solution of the EFA in the pre-testing.



Note. CFA = Confirmatory factor analysis; DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information

* $p < .05$; **$p < .01$; ***$p < .001$.

Figure 5-2 One-factor CFA-model of DAC

The results of the CFA for the one-factor model are presented in Table 5-12. Besides the Chi-square (χ^2) statistic, the model fit indices of CFI, RMSEA, and the Standardized Root Mean Square Residual (SRMR) were examined. For assessing the one-factor model displayed in Figure 5-2 the cut-off criteria of Brown and Cudeck (1993), and Hu and Bentler (1999) were used.

The Chi-square test statistics turned out to be significant[75] (χ^2 = 17.527, p = .041). However, the threshold for CFI ≥ .95 was met with CFI = .974, as well as the cut-off of RMSEA ≤ .08 with RMSEA = .051, and the benchmark for SRMR ≤ .08 with SRMR = .031. Referring to these three model fit indices, the proposed one-factor model shows an acceptable fit.

[75] To affirm a model fit, Chi-square should not become significant (e.g. Sörbom, 1975).

Table 5-12 CFA of the one-factor model for DAC

		Overall model fit				
Factor/construct	Indicator/dimension	χ^2	*df*	CFI	RMSEA	SRMR
DAC	II FD DU IA DF DO	17.527*	9	.974	.051[a]	.031

Note. N = 368; CFA = Confirmatory factor analysis; DAC = Decision-Analytical Competence; χ^2 = Chi-square; *df* = degrees of freedom, CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; II = Ability to integrate information; FD = Ability to frame a decision; DU = Ability to deal with uncertainty; IA = Ability to identify relevant alternatives; DF = Ability to assess decision fitness; DO = Ability to recognise decision opportunities.

* $p < .05$; ** $p < .01$; *** $p < .001$.

[a] 90 per cent confidence interval: .010 - .086.

Table 5-13 gives an overview of the factor loadings (*λ*), the intercepts, and the error variances for each of the six DAC-dimensions. All standardised factor loadings appear above $\lambda = .4$, ranging from $\lambda_{stand} = .421$ for OPPORTUNITIES to $\lambda_{stand} = .688$ for INFORMATION, and thereby fit the interpretation advice of Guadagnoli and Velicer (1988). There is, thus, reason to believe that the latent construct of DAC is able to significantly explain all DAC-dimensions.

Table 5-13 Factor loadings, intercepts, and variances of the one-factor CFA-model for DAC

	$\lambda_{unstand}$	*SE*	*z*-value	λ_{stand}	*p*
Loadings					
DF	0.758	.102	7.407	.488	.000
FD	1.018	.094	10.855	.587	.000
DU	0.968	.099	9.956	.568	.000
II	2.391	.171	13.972	.688	.000
DO	0.795	.120	6.601	.421	.000
IA	0.766	.088	8.657	.525	.000
Intercepts					
DF	5.443	.081	67.220	3.506	.000
FD	3.692	.091	40.737	2.131	.000
DU	4.041	.091	44.524	2.326	.000
II	6.639	.181	36.620	1.910	.000
DO	5.936	.099	59.744	3.144	.000
IA	6.109	.076	80.233	4.188	.000
Error Variances					
DF	1.835	.142	12.943	0.762	.000
FD	1.966	.179	11.000	0.655	.000
DU	2.046	.183	11.155	0.678	.000
II	6.367	.670	9.509	0.527	.000
DO	2.932	.218	13.444	0.823	.000
IA	1.541	.132	11.693	0.724	.000

Note. N = 368; CFA = Confirmatory factor analysis; DAC = Decision-analytical competence; $\lambda_{unstand}$ = unstandardised estimate for factor loadings; *SE* = standard error; λ_{stand} = standardised estimate for factor loadings; DF = Ability to assess decision fitness; FD = Ability to frame a decision; DU = Ability to deal with uncertainty; II = Ability to integrate information; DO = Ability to recognise decision opportunities; IA = Ability to identify relevant alternatives.

Overall, the DAC-test presents relatively good internal consistency. The *z* scores of its 49 items demonstrate a Cronbach's alpha of α = .85. The *z* scores of its six sum score values for OPPORTUNITIES, FITNESS, FRAME, ALTERNATIVES, UNCERTAINTY, and INFORMATION only marginally miss the desired threshold for Cronbach's alpha with α = .69, but meet the threshold for McDonald's omega by ω_H = .72.

5.2.4 Validity of the DAC-Test

After examining reliability, test score distribution, and dimensionality of the DAC-test, the current sub-section shows the results of the validation. As nomological validity expresses the degree of accuracy with which "the measure fits lawfully into a network of expected relationships" (Nunnally & Bernstein, 1994, p. 91), the correlations of the DAC-test score with a set of chosen manifest variables, so-called *validation criteria*, are presented. If the found correlations fit the theoretical interdependences as described in sub-sections 5.1.1 - 5.1.7, it can be assumed that the DAC-test results can be attributed to the to-be-measured latent construct.

Decision Self-Esteem

In the present research, decision self-esteem was measured by the self-rating Melbourne Decision Making Questionnaire I (Mann et al., 1998). As three of its six items are negatively poled, they had to be reversed for the correlation analysis. On a 3-point scale, decision self-esteem was assessed on average with 1.97 (*SD* = 0.65). The first row of Table 5-14 shows low positive correlations in the expected direction between decision self-esteem and DAC, i.e. for the DAC-dimensions and the overall scores. While OPPORTUNITIES, FRAME, ALTERNATIVES, UNCERTAINTY, and DAC-test overall are significantly related to decision self-esteem, FITNESS and INFORMATION are not. Apart from these two DAC-dimensions, DAC is significantly associated with decision self-esteem (cf. sub-section 5.1.1).

Table 5-14 Correlations between DAC-dimensions, validation criteria, and socio-demographic variables

	DO (+)	DF (+)	FD (+)	IA (+)	DU (+)	II (+)	DAC (+)
DSE (+)	.170**	.100	.114*	.146**	.171**	.083	.181***
V (+)	.119*	.097	-.004	.103*	.107*	.143*	.167*
B (-)	-.090	.044	-.049	-.055	-.028	.057	-.003
P (-)	-.107*	-.064	-.066	-.062	-.078	-.066	-.100
H (-)	-.059	.030	-.066	-.057	-.117*	-.092	-.096
SwD (+)	.087	.055	.078	.004	.057	.066	.098
IoU (-)	-.134*	-.127*	-.104*	-.144**	-.173***	-.127*	-.204***
HMT-S (+)	.238***	.238***	.270***	.319***	.378***	.467***	.488***
ToH (+)	.176**	.216***	.165**	.168**	.203***	.331***	.340***
Age (+)	.049	-.029	.064	-.059	-.006	.044	.026
GPA (+)	.068	.141**	.107*	.102	.151**	.091	.185***
PA-M (+)	.026	.154**	.135*	.148**	.314***	.212***	.279***

Note. The sign pictured next to each variable indicates whether higher scores reflect better (+) or worse (-) performance. DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information; DSE = Decision self-esteem (Melbourne Decision-Making Questionnaire I); V = Melbourne Decision-Making Questionnaire II - vigilance; B = Melbourne Decision-Making Questionnaire II – buck-passing; P = Melbourne Decision-Making Questionnaire II – procrastination; H = Melbourne Decision-Making Questionnaire II – hypervigilance; SwD = Satisfaction with Decision Scale; IoU = Intolerance of Uncertainty Scale; HTM-S = Hagen Matrices Test – Short version; ToH = Tower of Hanoi total moves; GPA = general point average; PA-M = point average in maths.

*Two-sided $p < .05$; **$p < .01$; ***$p < .001$.

Decision-Making Style

Test participants' style of decision-making was measured by the 22 items of the Melbourne Decision Making Questionnaire II (Mann et al., 1997) distinguishing between *vigilance* (*M* = 2.56, *SD* = 0.59), *buck-passing* (*M* = 1.71, *SD* = 0.71), *procrastination* (*M* = 1.62, *SD* = 0.69), and *hypervigilance* (*M* = 1.72, *SD* = 0.72) on its 3-point scale. Rows two to five of Table 5-14 present the correlation of these decision-making style dimensions and DAC. While all DAC-test sub-scales but FITNESS and FRAME, show significant positive correlations with the effective decision-making style *vigilance*, the results for the three remaining less constructive decision-making styles, i.e. *buck-passing*, *procrastination*, and *hypervigilance*, are either negative and/or not significant. According to *vigilance,* the results of DAC's validation appear in the expected direction (cf. sub-section 5.1.2).

Decision Satisfaction

On a 5-point scale decision satisfaction was measured by the 6-item questionnaire of Holmes-Rovner and colleagues (1996) named Satisfaction with Decision Scale. Test participants evaluated their decision satisfaction on average at 3.95 (*SD* = 0.82). In Table 5-14 the sixth row presents the slight positive but not significant correlations between the DAC components and decision satisfaction. The results appear in the expected direction but they are not significant.

Intolerance of Uncertainty

The tendency of people's inability to act in cases of perceived uncertainty was measured by the correspondent sub-scale of the Intolerance of Uncertainty Scale of Buhr and Dugas (2002). Mean self-ratings of 2.43 (*SD* = 1.19) were produced by the negatively-keyed scale. The correlations of the DAC components and the IoU are shown in row seven of Table 5-14. As expected, all DAC-dimensions and the overall DAC-test score correlate significantly negative with the inactivity due to perceived uncertainty. Of the DAC-dimensions, UNCERTAINTY presents the highest negative association with the self-assessed inability to act in cases of perceived uncertainty (cf. sub-section 5.1.4).

Fluid Intelligence

Fluid intelligence was operationalized by the short version of the Hagen Matrices Test of Heydasch et al. (2013) consisting of six matrices. The number of correctly solved matrices measures the performance of this scale. Of all test participants, 244 (66.12%) completed the HMT-S. On average, test participants solved 4.41 matrices (*SD* = 1.46). Row eight of Table 5-14 gives an overview of the significant moderate to good correlations with DAC. With OPPORTUNITIES, FITNESS, and FRAME the HMT-S shows moderate positive correlations. With ALTERNATIVES, UNCERTAINTY, INFORMATION, and the

DAC overall score, the HMT-S presents good positive correlations. Thus, it can be assumed that DAC is associated with fluid intelligence (cf. sub-section 5.1.5).

Problem-Solving Competence

The Tower of Hanoi with four disks/rings served to measure problem-solving competence in the present research. Performance was calculated by the total number of *moves* test participants needed to complete the task. Two hundred and seventy-nine test participants (75.61%) completed the ToH. Of those, test participants needed on average 25 moves ($M = 24.71$; $SD = 10.14$) to solve the problem-solving task. As the fewer moves the better, the scale is negatively-keyed and had to be reversed for the validation analysis. Row nine of Table 5-14 shows the correlation with DAC. As presumed problem-solving competence correlates positively with all DAC components (cf. sub-section 5.1.6), these significant correlations support the validation of DAC.

Socio-Demographic Variables

Row 10 of Table 5-14 presents the correlation of the DAC components with age. The results show slight positive and negative, but, as assumed, no significant relations.

The calculated correlations of the DAC components with the GPA are shown in row 11 of Table 5-14. For all DAC components the relationships appears to be positive. While values for FITNESS, FRAME, UNCERTAINTY, and the overall DAC-score are significant, values for the remaining DAC-dimensions are not.

Row 12 of Table 5-14 gives an overview of the found correlations between the DAC components and test participants' point average in maths. As presumed, relationships are positive (cf. sub-section 5.1.7), significant for all DAC components except OPPORTUNITIES. The highest correlation can be found between point average in maths and UNCERTAINTY – a DAC-dimension, which relies on numeracy.

5.2.5 Objectivity of the DAC-Test

Besides reliability and validity, objectivity belongs to the three intended key quality criteria for psychometric measures. According to this important aim in the process of test development, the following short sub-section argues to what extent the test results concerning testing procedure, the test analysis, and the results interpretation can be assumed to be independent of the researcher.

Referring to the testing procedure, the DAC-test was exclusively executed online. The order of task appearance was kept the same including its display of tasks and items for all test participants. The complete set of received digital data was automatically saved on a server and within this process immediately coded. The set of closed items of the DAC-test led to comparable and analysable data. As the performance test contained concrete and precise definitions of the *correct* answer options for each item, the analysis of results did not provide any tolerance for interpretations. All data analyses were run with the same R-scripts or SPSS syntaxes.

Concluding, objectivity was paid attention to for the test development. Given the DAC-test's circumstances described above, it is assumed that the results of the test are independent of the researcher.

After the main testing results of this PhD research have been presented in the current chapter, chapter 6 discusses these results by highlighting ways of explanation, by revising the hypotheses, by comparing the DAC-test with the A-DMC performance test of Bruine de Bruin and colleagues (2007), and by pointing out the limitation and potential for further research of the present avenue of research.

6 RESULTS DISCUSSION

The discipline of decision analysis allows a problem to be broken down into comprehensible components (Edwards, Miles Jr., et al., 2007) and thereby provides a systematic procedure that supports people in making sound decisions and sorting out problems (Bazerman & Moore, 2009). The main aim of the present research was to operationalize the skill of analytical decision-making by an objective, reliable, and valid psychometric performance test. The developed DAC-test can be understood as an achievement test for decision-analytical knowledge. Consequently, DA is not seen as an aptitude. The DAC-test measures how well a person is able to follow decision-analytical reasoning steps to solve a complex decision problem. While chapter 4 presented the operationalization of the construct and the results of the pre-testing, the preceding chapter 5 constituted the results of the main testing according to reliability and validity of the DAC-test. The current chapter carefully considers these results and offers interpretations.

Thereto, section 6.1 discusses findings and presents attempts to explain certain results. Section 6.2 answers the to-be-examined hypotheses, while section 6.3 presents a summarising contrasting juxtaposition of the A-DMC index (Bruine de Bruin et al., 2007) and the developed DAC-test by their quality criteria. Scientific contributions and potential practical utilisation are addressed by section 6.4. Limitations of the findings and directions for further research close this chapter in section 6.5.

6.1 DISCUSSION OF THE FINDINGS & ATTEMPTS TO EXPLAIN

The presented results of the empirical control of this Ph.D. research (cf. section 5.2) are partially strong in evidence and thus distinct to interpret. However, several findings are not that explicit and as a result do not satisfy the expectations and set goals for this Ph.D. research. For this reason, the subsequent sub-sections discuss the results and highlight various approaches of explanation.

The first sub-section 6.1.1 presents reflections on reliability and dimensionality of the DAC-test on dimensional level and sub-section 6.1.2 of the DAC-test as a whole. In sub-section 6.1.3, reflections on validity of the aggregated DAC-test score are discussed.

6.1.1 Reflections on Reliability & Dimensionality of the DAC-Dimensions

Derived from decision-analytical literature, eight cognitive dimensions (cf. sub-sections 2.4.1 to 2.4.3) have been identified to capture DAC, of which six were successfully operationalized by a performance-based set of decision tasks. Following this, the item sets of the DAC-dimensions OPPORTUNITIES, FITNESS, FRAME, ALTERNATIVES, UNCERTAINTY, and INFORMATION were tested in the pre-testing. According to the thresholds for the two quality criteria of the item analysis[76], items with an item discriminatory power of $r_{it} < .2$ (cf. Fisseni, 2004) and/or an item difficulty index of $p_i > .8$ and $p_i < .2$ (Kline, 2015, p. 143; Lienert & Raatz, 1998, p. 115) were excluded from further analyses and deleted from the test.

In the main testing, the split-half analyses led to the exclusion of four more items. Thus, the dimensions OPPORTUNITIES, FITNESS, ALTERNATIVES, and UNCERTAINTY each retained one item less than in the pre-testing. As appeared in the pre-testing, only INFORMATION reached the desired cut-off of .70 for internal consistency measured by Cronbach's alpha (e.g. Bland & Altman, 1997; Nunnally et al., 1978). However, INFORMATION is the only sub-scale that consists of more than 10 items. As the internal consistency coefficient reacts sensitive to the number of items considered, but not so extreme for scales with more than 10 items (R. A. Peterson, 1994), the length of the INFORMATION scale could have caused the high Cronbach's alpha. While the coefficients for FITNESS ($\alpha = .33 \rightarrow \alpha = .47$) and ALTERNATIVES ($\alpha = .44 \rightarrow \alpha = .51$) showed an increase, the values for OPPORTUNITIES ($\alpha = .63 \rightarrow \alpha = .58$), FRAME ($\alpha = .51 \rightarrow \alpha = .42$), and UNCERTAINTY ($\alpha = .64 \rightarrow \alpha = .61$) decreased from the pre-testing to the main testing. In the cases of OPPORTUNITIES and UNCERTAINTY the reduction of items could be assumed as one reason for the decrease (Nunnally et al., 1978). But still, other arguments as presented in sub-section 4.2.5, have to be taken into account.

Multi-dimensionality of the underlying constructs could be considered as another reason for low Cronbach's alphas (Schmitt, 1996). Subsequently, in order to analyse the dimensionality a CFA for each DAC-dimension was run. Values below 1 for the WRMR confirmed the homogeneity for the scales of FITNESS, FRAME, and ALTERNATIVES. The scales of OPPORTUNITIES, UNCERTAINTY, and INFORMATION appeared to be heterogeneous (cf. Table 5-9). The inspection of the referring factor loadings of exploratory factor analyses for these three DAC-dimensions did not offer valuable clues to the underlying structure. Thus, the dimensionality of these DAC sub-scales could not exclusively be resolved.

One suspicion concerning the low internal consistencies of OPPORTUNITIES, FITNESS, FRAME, ALTERNATIVES, and UNCERTAINTY and the detected unintentionally negatively-poled items are contradictory answers across test participants within the said DAC-dimensions.

[76] Appendix sub-section 9.12 displays item difficulty indices and discriminatory power coefficients for each DAC-dimension of the main testing.

So, it is presumable that not all test participants have solved all items in a consistent manner, i.e. a few participants might have randomly guessed to complete the test; either in case of some items or even in case of the whole set of items. As random guessing usually leads to a correlation close to r = 0 (cf. Bortz & Döring, 2005), a different effect might have appeared here: All four items, which had to be removed from the final analyses in the main testing were items with true/false response sets (cf. Table 6-1). Assuming that the DAC-test was "clicked through" by a certain group of test participants and not thoughroughly answered, the way of clicking might have had a systematic effect on the results. So, if this group has completed the test by clicking *enter – forward – enter – forward – enter* and so forth, a specific and maybe recurring answering-pattern might have occurred for this group of people[77]. In case of true/false items this potential pattern might have resulted in the finding that some items were somewhat negatively-answered.

Table 6-1 Tested DAC-dimensions, assessment criterion, and response scales

DAC-dimension	Required decision-making ability	Assessment criterion	Response scale
DO	Ability to identify decision situations, which are characterised by the potential to support decision-makers in achieving their desired goals, and to assess which degree of analytical thinking is needed to solve those situations.	Accuracy	a) True/False b) MC[a] with 3 options
DF	Ability to decide whether emotions and physical status allow focusing on a concrete decision situation, proceeding analytically with the decision-making process, and making a conscious choice.	Accuracy	True/False
FD	Ability to perceive the relevant aspects of a decision situation and to interpret their meaning for the corresponding process of decision-making.	Accuracy	MC[a] with 10 options
IA	Ability to envision alternatives that fit a set of given objectives in a decision situation.	Accuracy	True/False
DU	Ability to comprehend, interpret, and calculate probabilities of different qualities in order to make a choice.	Accuracy	a) True/False[a] b) MC[b] with 5 options c) MC[b] with 19 options d) Open format
II	Ability to combine assessments of uncertainties and values coherently, compare various alternatives and objectives, and correspondingly rank those alternatives in relation to their relevance for the decision-maker.	Accuracy	MC[b] with 7 options

Note. DAC = Decision-analytical competence; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.

[77] The position of the correct answer – either true or false – was varied within the scales.

[a] The only item of DU (DU_01) with a true/false response set was removed for final analyses.

[b] MC = Multiple-choice.

As Cronbach's alpha is restricted due to the various item loadings, the internal consistency coefficient named McDonald's omega, for which item loadings do not have to be equal, was additionally calculated. Three of the six DAC-dimensions met the desired threshold of McDonald's omega by $\omega_H \geq .70$ (Zinbarg, Revelle, Yovel, & Li, 2005): OPPORTUNITIES (ω_H = .76), UNCERTAINTY (ω_H = .84), and INFORMATION (ω_H = .97). While ALTERNATIVES (ω_H = .69) and FITNESS (ω_H = .66) barely missed the targeted value, FRAME (ω_H = .56) showed the lowest McDonald's omega. Deductively, according to McDonald's omega three of the six DAC-dimensions show good reliability by internal consistency.

To further analyse the reliability of each DAC-dimension, other reliability tests have to be applied (e.g. Schmitt, 1996). For the present research though, neither re-test reliability nor parallel-test reliability were applicable. For re-test reliability, the same group of people has to complete the test twice. For analysing the test's reliability, the test scores of the first and the second measuring have to be compared directly, i.e. the two test scores of each participant have to be assigned distinctly to each other. Due to the ensured confidentiality of this Ph.D. research project, such a case-wise comparison was not feasible.

Examining parallel-test reliability was the intention at the beginning of this research project. However, due to test economy and that test participants should be able to complete the test in a reasonable time, the originally planned second case - INVEST case (cf. sub-section 4.2.1) - had to be neglected.

6.1.2 Reflections on Reliability & Dimensionality of the DAC-Test

According to the proposal of researchers such as e.g. Stanovich and West (2000) or hitherto existing research of e.g. Bruine de Bruin and colleagues (2007), DMC can be assumed as a latent construct underlying various decision-making tasks. The present research intended to examine the internal consistency and dimensionality of the proposed DAC-test as a whole, containing six cognitive sub-scales with corresponding item sets. Therefore, in the pre-testing an EFA with varimax rotation was run. One factor was extracted and explained 29% of the observed variance. All factor loadings exceeded the recommended benchmark for interpretation of factor loadings with $\lambda \geq .30$ (Hair et al., 1998, p. 112). In the main testing, a CFA was applied to analyse the model fit of the one-factor structure. Its results presented a good overall model fit (cf. Hu & Bentler, 1999). Cronbach's alpha for the final set of 49 DAC items met the desired threshold and performed at α = .85 (Bland & Altman, 1997; Nunnally et al., 1978; Schmitt, 1996). The internal consistency of the six sub-scale sum scores reached α = .69 and ω_H = .72 forming one general factor.

A CFA was run to examine whether the concept of DAC fits the assumed structural composition of the concept presented in Figure 5-2. The main aim of the statistical procedure was to determine whether the gathered data is consistent with the hypothesised measurement model, which was based on preceding analytic research, ran by an EFA (cf. sub-section 4.2.4). While EFA is a structure-searching analytical tool, the CFA can only verify the expected pattern of a construct. Thus, the CFA cannot reverse theoretically-made decisions. Also, it is important to keep in mind that the various model fit indices do not confirm the correctness of the model, they only indicate the plausibility of the model (Schermelleh-Engel, Moosbrugger, & Müller, 2003). In this context, it is also important to point out that reliability measures turn out to be better when a more precise factor model is used.

The one-factor model met the cut-off criteria for CFI, RMSEA, and SRMR of Brown and Cudeck (1993) and Hu and Bentler (1999) in the CFA and the good internal consistency of the complete DAC-test item set supported the assumption of DAC being an explanatory latent factor explaining significant variance of the six sets of DAC-tasks; however, the missed threshold for Cronbach's alpha of the six DAC sub-scale sum scores gives reason to think about potential explanations, including that having six variables to calculate Cronbach's alpha affected the small internal consistency coefficient (Nunnally et al., 1978); also, the construct of DAC is too heterogeneous to show internal consistency (Schmitt, 1996), and some test participants might not have answered the DAC-test consistently and/or conscientiously and hence show a higher vulnerability to guessing. Table 6-1 gives an overview of the six tested DAC-dimensions including their required skills, the corresponding criterion for assessment, and the presented response scale. The three scales for OPPORTUNITIES, FITNESS, and ALTERNATIVES mainly go along with a true/false response set and hence show a vulnerability to guessing (Kline, 2015), i.e. the binary-scaled response sets involve a higher chance to correctly guess the answer. The remaining three sub-scales FRAME, UNCERTAINTY, and INFORMATION are predominantly presented as multiple-choice, i.e. an interval-scaled, response set.

One remark that has to be made is that the presented conclusions on DAC's dimensionality just refer to the six DAC-dimensions, which have been successfully operationalized within the framework of this research. As long as missing aspects of DAC, such as OBJECTIVES, have not been effectively surveyed and together with the other sub-scales examined, then no final statement about the construct's dimensionality can be made.

Although the DAC-test's structure was proven to coincide with theory on an aggregated level, some subscales displayed discrepancies. Further investigations are necessary to research the dimensionality of the DAC sub-scales and the construct as a whole.

6.1.3 Reflections on Validity of the DAC-Test

The validity of the DAC-test was examined for its nomological validity, which describes the accuracy with which the empirical results of the test conform to a set of theoretically founded relationships with other constructs (e.g. Nunnally & Bernstein, 1994). In this framework, the DAC-test was presented with a catalogue of 12 psychological scales and demographic variables, which served as validation criteria.

In general, the observed relationships between the DAC-test and its validation criteria displayed in Table 5-14 appeared in the expected directions (cf. sub-sections 5.1.1 - 5.1.7). Nevertheless, not all results were significant and especially correlations with self-assessment scales turned out to be lower than expected. According to the assumption that some test participants answered the DAC-test and its validation criteria in a non-consistent manner, e.g. by "clicking-through", less distinctive results are comprehensible. Also, it appears to be challenging to operationalize DAC holistically. The successful operationalization of two dimensions did not work out in the framework of the present research. In addition, the dimensionality of the DAC sub-scales needs further investigation. So far, DAC seems to be a complex construct; theoretically homogeneous and empirically calling for more insights. Linking the further development of both theory and measurement should thus be more pronounced.

With decision self-esteem measured by the Melbourne Decision Making Questionnaire I of Mann et al. (1998) the overall DAC-test score showed a significant relationship of r = .181 (p = .001). A vigilant style of decision-making and DAC-test demonstrated a significant positive correlation of r = .167 (p = .05). The less respectable styles of decision-making, i.e. buck-passing, procrastination, and hypervigilance, demonstrated negative but not significant correlations with DAC. All four styles of decision-making were captured by the Melbourne Decision Making Questionnaire II of Mann et al. (1997). The empirical results indicated that higher DAC-test scores are associated with lower scores of the sub-scale *uncertainty leads to the inability to act* (r = -.204; p = .001) of the Intolerance of Uncertainty Scales of Buhr & Dugas (2002). The scores of decision satisfaction (Satisfaction with Decision Scales by Holmes-Rovner et al., 1996) were slightly positive but not significantly linked to higher DAC-test scores.

In contrast to the findings of these self-rating measures, the results for both performance-based measures, for fluid intelligence and problem-solving competence, were more distinctive. So, DAC and fluid intelligence, the latter construct measured by the short version of the Hagen Matrices Test (HMT-S; Heydasch, Haubrich, & Renner, 2013), demonstrated a good positive relationship of r = .488 (p = .001). Better performance in problem-solving competence, assessed by the Tower of Hanoi (ToH; for instance Anderson & Douglass, 2001; H. J. Bell, 2004; Schiff & Vakil, 2015; Vernon & Strudensky, 1988), was related significantly to higher scores of the DAC-test (r = .340; p = .001).

One possible explanation for the difference in the distinctiveness of the correlation between the DAC-test and self-rating vs. performance-based measures, might be the self-selection bias (e.g. Hudson, Seah, Hite, & Haab, 2004; Whitehead, 1991). The self-selection bias appears in any situation where test participants assign themselves to a specific group. In case of the *Tower of Hanoi* task, test participants were able to quit the task. This opportunity was given to allow test participants who were not able to solve the task to continue with the survey. Due to this self-selection a certain group, 75.61% of test participants, skipped the task – some could not successfully complete the task; some were not motivated enough. To examine this conjecture, a logistic regression for binary data was run. It shows that the probability of completing the ToH increases with rising overall DAC-test scores. Indeed, the chance to accomplish the ToH duplicates with each standard deviation of the DAC-test ($e^x(0.68) = 1.97$).

In a comparable style, the results of the HMT-S have to be considered: For administrative reasons, Heydasch and colleagues did not agree to embed their matrices test into the existing DAC-test website. For this reason, the HMT-S was appended at the very end of the survey, following the DAC-test, the catalogue of validation criteria, and the socio-demographic variables. Sixty-six per cent of all test participants who completed the DAC-test also submitted the HMTS-S. Consequently, the self-selection bias might also have occurred in this case. The more motivated or competent test participants may have completed both the DAC-test and the HMT-S, and therefore produced the scores that form the correlation results. However, here the logistic regression showed that the final DAC-test score is not significantly correlated (p = .194) with the probability of completing the HMT-S.

In terms of the non-significant relationship of DAC and decision satisfaction it has to be noted that the SwD in its original study (Holmes-Rovner et al., 1996) was presented in the context of acute postmenopausal hormone-replacement therapy decisions. In comparison to the present study, test participants were able to relate their self-assessment to one concrete real life decision. In this Ph.D. research, test participants were asked to envision the last complex decision they had to deal with. Consequently, test results might be very heterogeneous in terms of how long the considered decision lies in the past or whether output and/or outcome for this individual decision were already known. This implied heterogeneity might be one reason why the scores of DAC and the SwD do not show a significant relationship.

"While the need to measure post-decision satisfaction has been recognized in previous studies, the resulting measurements have been limited, both in scope and in depth." (Sainfort & Booske, 2000, p. 53) In comparison to other scales, such as Houston, Sherman, and Baker's one-item measure to assessing how happy persons are with their choice (1991) or Jones' 32-item *Career Decision Profile* (1989), the SwD of Holmes-Rovner et al. (1996) was chosen as it presented a good trade-off between length of the scale with its six items and content fit capturing, among other things, the consistency with personal objectives or the comparison of alternatives. Due to the complexity in terms of content and duration of the DAC-test and its catalogue of validation criteria, even scales with an average of 10 items, as for instance the *Decision Making Quality Scale* of Hollen (1994) were neglected.

Between age and DAC performance no significant relationship could be observed. According to the theoretical expectations, the age range of the target group might have been too tight, as test participants between 18 and 30 years belong to the same phase of psychological development (cf. Erikson, 1959).

As found by the correlation of the DAC-test with the HMT-S, the tasks of the DAC-test seem to capture fluid intelligence. That may be one explanation why the DAC-test and the GPA significantly positively correlate. Also, the average mark in maths, which relies on numeracy, shows a significant positive relationship with the overall DAC-test score and especially with UNCERTAINTY – a scale that mainly captures mathematical skills.

Some initial evidence about the validity of DAC as a psychological construct could be observed by the present study; validated by decision-self esteem, vigilant decision-making style, intolerance of uncertainty, fluid intelligence, problem-solving competence, GPA and average mark in maths.

The following section addresses the achievement of the set research hypotheses.

6.2 REVISION OF RESEARCH HYPOTHESES

The statistical-methodological research hypothesis H_1 claims that each dimension of the DAC-test is reliable and thus shows appropriate internal consistency. To measure internal consistency on a sub-scale level two different coefficients were calculated: Cronbach's alpha (Cronbach, 1951) with its threshold of $\alpha \geq .70$ (e.g. Nunnally et al., 1978) and McDonald's omega (McDonald, 1999) with its desired value of $\omega_H \geq .70$ (Zinbarg et al., 2005). Other reliability coefficients could not be applied. So, calculating re-test reliability was not possible due to vested confidentiality[78] and missing research resources such as budget. Parallel-test reliability was intended at the beginning of this research. However, due to test economic issues the designed parallel form of the EDU case of the DAC-test, the INVEST case, could not be tested in the main testing.

According to the internal consistency coefficients of Cronbach's alpha and McDonald's omega, Hypothesis H_1 could only partially be verified. While the threshold for Cronbach's alpha was only met by INFORMATION (α = .86), the DAC-test sub-scale with more than 10 items, the threshold for McDonald's omega was met by OPPORTUNITIES (ω = .76) UNCERTAINTY (ω = .84), and INFORMATION (ω = .97). Thus, it can be stated that three DAC-test sub-scales show internal consistency and are thus reliable measures.

Research Hypothesis H_2, which refers to the reliability of the DAC-test as a whole, could be verified in terms of the internal consistency for the final set of 49 DAC-test items, which showed a Cronbach's alpha of α = .85. Cronbach's alpha of the six *z*-scores of the sum score values for OPPORTUNITIES, FITNESS, FRAME, ALTERNATIVES, UNCERTAINTY, and INFORMATION is α = .69. However, McDonald's omega for the z-scores of the DAC-test sum score values meets the threshold and performs at ω = .72.

The third statistical-methodological research hypothesis H_3, which states that the DAC-dimensions refer to one general factor of decision-making, could be verified. In the pre-testing, an EFA presented a one-factor solution for the *z*-scores of the six analysed

[78] The self-created identification code of each test participant (cf. appendix section 9.10 first paragraph) was automatically saved in the database after test completion. It served to reconstruct who of the test participants had successfully finished the testing and deserved the remuneration of €15.00. However, the identification code was saved in another data sheet, independent of the actual test data. An assignment of test and re-test data would not be possible.

DAC-dimensions with factor loadings above λ = .30 (Hair et al., 1998, p. 112). It captured 29% of the variance. The CFA of the main testing confirmed a good model fit for the tested one-factor structure of the six *z*-scores of the DAC-dimensions. Even though chi-square was significant (Sörbom, 1975; χ^2 = 17.527, *p* = .041), the thresholds by Hu and Bentler (1999) for the three model fit indices CFI ≥ .95, RMSEA ≤ .08, and SRMR ≤ .08 were met with CFI = .974, RMSEA = .051, and SRMR = .031.

The content-valid research hypothesis H_4, which claims that the DAC-test score correlates significantly with suitable criteria for validation, is partially seen as verified. To validate the DAC-test, correlations with a catalogue of 12 validation criteria were calculated. The DAC-test showed positive significant relationships with decision self-esteem (r_P = .181, p < .001), vigilant decision-making style (r_P = .167, p < .05), fluid intelligence (r_P = .488, p < .001), problem-solving competence (r_P = .340, p < .001), the GPA (r_P = .185, p < .001), and the average mark in maths (r_P = .279, p < .001). The DAC-test presented a negative significant correlation with the inability to act in case of perceived uncertainty, a sub-scale of intolerance of uncertainty (r_P = -.204, p < .001). All correlations appeared in the expected direction. However, the correlations of DAC with decision satisfaction, buck-passing, procrastination, and hypervigilance turned out to be not significant.

Summarising, the majority of the research hypotheses of this Ph.D. research could be verified. Reliability in terms of internal consistency measured by McDonald's omega could be proven for three DAC-dimensions, OPPORTUNITY, UNCERTAINTY, and INFORMATION, as well as on test-level. The assumption of DAC being a higher-order factor underlying various dimensions was supported by the results of the EFA and CFA. Evidence for the DAC-test's validity was found, for instance, by significant positive correlations of DAC and decision self-esteem, vigilance, fluid intelligence, or problem-solving competence.

The subsequent section compares the A-DMC of Bruine de Bruin et al. (2007) and the developed DAC-test of the present research on content criteria and statistical properties.

6.3 COMPARISON OF THE A-DMC & THE DAC-TEST

In the present section, the developed DAC-test is compared to so far one of the most promising approaches of measuring DMC by a performance test – the A-DMC of

Bruine de Bruin and colleagues (2007). The tables below serve to give a compact overview of the two measures by their data basis and statistical properties.

While the A-DMC is based on the normative approach to decision-making, the DAC-test aims to measure prescriptive DMC (cf. Table 6-2). Hence, the A-DMC concentrates on behavioural decision theory, whereas the DAC-test focuses on analytical decision theory. The target group is set much broader for the A-DMC than for the DAC-test. The A-DMC is designed for adults in general. This fact becomes apparent by the age range of the sample group in the study of Bruine de Bruin and colleagues (2007). Here, the mean age of the 360 study participants, which were recruited by social service organisations and community groups, lies at 48 years (*M* = 47.7; *SD* = 17.0) with a range of 18 to 88 years. Seventy-three point eight per cent of the A-DMC study participants were female and 26.2% male. In comparison, young adults between 18 and 30 years built the target group for the DAC-test. The corresponding 368 study participants were acquired via universities and research institutes. The average age of study participants was 24 years (*M* = 23.8; *SD* = 3.4), of which 52.5% were women and 47.5% men.

Table 6-2 Comparison of the A-DMC and the DAC-test by theoretical basis, target –, and sample group

Criterion	Specification	A-DMC	DAC-test
Theoretical basis		Normative approach to decision-making[79]	Prescriptive approach to decision-making
Target group		Adults	Young adults
Sample group	***N***	*N* = 360	*N* = 368
	Group specifics	Acquired through social service organisations and community groups	Acquired through universities and research institutes
	Age	*M* = 47.7 *SD* = 17.0 Range = 18-88	*M* = 23.8 *SD* = 3.4 Range = 18-30
	Sex	Female: 73.8% Male: 26.2%	Female: 52.5% Male: 47.5%

Note. A-DMC = Adult Decision-Making Competence index score (Bruine de Bruin et al., 2007); DAC-test = Decision-Analytical Competence test; *N* = Total number of cases.

Table 6-3 gives an overview of the covered dimensions of both tests and presents the dimensions' statistical properties. The A-DMC captures seven dimensions: resistance to framing, recognising social norms, under/overconfidence, applying decision rules,

[79] Bruine de Bruin et al. (2007) describe that the A-DMC refers to the normative approach of decision science. However, in comparison to the DAC-test, it could be stated that both tests, the A-DMC and the DAC-test, intend to measure deviations from normative decision-making standards, while the A-DMC covers behavioural decision-making aspects and the DAC-test captures decision-analytical aspects.

consistency in risk perception, resistance to sunk costs, and path independence. The dimensions are represented each by a set of six (*path independence*) to 34 (*under/overconfidence*) items. Four of the seven dimensions meet the threshold for Cronbach's alpha of α = .70 (e.g. Nunnally et al., 1978). The re-test reliability of the A-DMC sub-scale lies between r_{tt} = .28 (p < .001) for path independence and r_{tt} = .77 (p < .001) for applying decision rules. Only one of the seven A-DMC dimensions meets the desired values for the re-test reliability of r_{tt} = .70 (Guilford, 1956).

Six cognitive dimensions operationalize DAC: ability to recognise a decision opportunity, ability to assess decision fitness, ability to frame a decision, ability to identify relevant alternatives, ability to deal with uncertainty, and ability to integrate information. Of the six sub-scales, one meets the threshold for Cronbach's alpha. However, the item sets of the DAC-test consist on average only of half as many items as the sub-scales of the A-DMC, varying between seven and 12 items per dimension. In terms of McDonald's alpha, three DAC sub-scales meet the threshold for internal consistency by ω_H > .70.

Table 6-3 Comparison of the A-DMC and the DAC-test by their dimensions and the corresponding statistical properties

Criterion/ statistical property	Specification	A-DMC	DAC-test
Dimensions	**Dimensions**	1) Resistance to framing 2) Recognising social norms 3) Under/overconfidence 4) Applying decision rules 5) Consistency in risk perception 6) Resistance to sunk costs 7) Path independence	1) Ability to recognise a decision opportunity 2) Ability to assess decision fitness 3) Ability to frame a decision 4) Ability to identify relevant alternatives 5) Ability to deal with uncertainty 6) Ability to integrate information
	Number of items; *Cronbach's alpha;* re-test reliability; McDonald's omega	1) 14; α = .62; r_{tt} = .58***; -- 2) 16; α = .64; r_{tt} = .46***; -- 3) 34; α = .77; r_{tt} = .47***; -- 4) 13; α = .73; r_{tt} = .77***; -- 5) 10; α = .72; r_{tt} = .51***; -- 6) 10; α = .54; r_{tt} = .61***; -- 7) 6; α = .75; r_{tt} = .28***; --	1) 8; α = .58; --; ω_H = .76 2) 7; α = .47; --; ω_H = .66 3) 8; α = .42; --; ω_H = .56 4) 7; α = .51; --; ω_H = .69 5) 7; α = .61; --; ω_H = .84 6) 12; α = .86; --; ω_H = .97

Note. A-DMC = Adult Decision-Making Competence index score (Bruine de Bruin et al., 2007); DAC-test = Decision-Analytical Competence test; α = Cronbach's alpha; r_{tt} = re-test reliability (p < .001); ω_H = McDonald's omega.

* *p* < .05; ***p* < .01; ****p* < .001.

Criteria for the internal consistency on construct-level and dimensionality are shown in Table 6-4. While the 103 items of the A-DMC present a Cronbach's alpha of α = .85, the 49 items of the DAC-test demonstrate the same value for internal consistency. The seven A-DMC *z*-scores of the dimensional sum scores show Cronbach's alpha at α = .83. The six *z*-scores of the DAC-dimension's sum scores present an internal consistency at α = .69 and ω_H = .72. For examining the dimensionality of the construct, an EFA with non-orthogonal oblimin rotation was used for the A-DMC. The one-factor solution explained 30% of the variance, whereby two of the seven sub-scales miss the threshold for factor loadings of $\lambda \geq .30$ stated by Hair et al. (1998, p. 112). For the DAC-test, the EFA with non-orthogonal promax rotation presents a one-factor solution, which explains 29% of the variance. All sub-scales met the threshold for factor loadings. Additionally, to verify the theoretical assumptions that general DMC builds a latent construct underlying various correlated dimensions (e.g. Stanovich & West, 2000), for the DAC-test the CFA presented a good fit for the one-factor model. For the A-DMC, a CFA to confirm the one-factor solution was not specified.

Table 6-4 Comparison of the A-DMC and the DAC-test by internal consistency and dimensionality

Criterion/ statistical property	Specification	A-DMC	DAC-test
Internal consistency...	...of all items	α = .85	α = .85
	...of the *z*-scores of the test's dimensions	α = .83	α = .69; ω_H = .72
EFA	Rotation	Oblimin rotation → non-orthogonal	Promax rotation → non-orthogonal
	Number of extracted factors	One[80]	One
	Factor loadings	1) λ = .48 2) λ = .40 3) λ = .35 4) λ = .80 5) λ = .49 6) λ = .23 7) λ = .10	1) λ = .45 2) λ = .36 3) λ = .57 4) λ = .55 5) λ = .62 6) λ = .66
	Percentage of explained variance	30%	29%
CFA	Tested model	--	One-factor model
	Model fit and fit indices	--	Good: CFI = .974 RMSEA = .051 SRMR = .031

Note. A-DMC = Adult Decision-Making Competence index score (Bruine de Bruin et al., 2007); DAC-test = Decision-Analytical Competence test; EFA = Exploratory factor analysis; CFA = Confirmatory factor analysis; α = Cronbach's alpha; ω_H = McDonald's omega; λ = Factor loadings; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table 6-5 juxtaposes the validation criteria for both tests and the corresponding correlations with significance level. To validate the A-DMC, nine self-rating scales, two performance measures, and three socio-demographic variables were co-tested. The results show positive correlations of A-DMC with the decision-making style using behavioural coping (r_P = .28; p < .001) and deciding rationally (r_P = .22; p < .001). Negative relationships are found with the decision-making style of feeling regret (r = -.14;

[80] According to Bruine de Bruin et al., (2007), the EFA actually extracted a two-factor solution. The authors argue that the loadings of the two-factor model "resemble those of the one-factor solution. Recognizing Social Norms, Resistance to Sunk Costs, and Path Independence have a higher loading on the second factor, but the latter remains under .30. The two-factor solution does not correspond to the three-factor solution reported for the Y-DMC (Parker & Fischhoff, 2005) [the prototype of the A-DMC]. Nor does either factor solution correspond to any of the three task characteristics ... : response mode, criterion, or general decision-making skills." (Bruine de Bruin et al., 2007, p. 944)

$p < .05$), needlessly maximising ($r_P = -.19$; $p < .001$), avoiding decisions ($r_P = -.21$; $p < .001$), and deciding spontaneously ($r_P = -.29$; $p < .001$). The A-DMC also correlates with the DOI ($r_P = .29$; $p < .001$) – the Decision Outcome Inventory - an introduced "self-report measure of decision-making success in terms of avoiding negative decision outcomes." (Bruine de Bruin et al., 2007, p. 943) The strongest correlations appeared with nonverbal cognitive ability ($r_P = .61$; $p < .001$), crystallised verbal ability ($r_P = .50$; $p < .001$), using social services ($r_P = -.54$; $p < .001$), and edcuation ($r_P = .47$; $p < .001$).

Table 6-5 Comparison of the A-DMC and the DAC-test by validity

Criterion/ statistical property	Specification	A-DMC	DAC-test
Validation criteria	Self-rating scales	**Decision-making styles** 1) To feel regret 2) To needlessly maximize 3) To use behavioural coping 4) To decide rationally 5) To decide intuitively 6) To depend on others 7) To avoid decisions 8) To decide spontaneously **Experienced decision outcomes** 9) DOI	**Decision-making styles** 1) Decision self-esteem 2) Vigilance 3) Buck-passing 4) Procrastination 5) Hypervigilance 6) Decision satisfaction 7) Intolerance of uncertainty
	Performance measures	**Cognitive abilities** 10) Nonverbal cognitive ability 11) Crystallised verbal ability	**Cognitive abilities** 8) Fluid intelligence 9) Problem-solving competence
	Socio-demographics	12) Using social services 13) Education 14) Age	10) Age 11) GPA 12) PA-M
	Correlations & level of significance	1) r_p = -.14* 2) r_p = -.19*** 3) r_p = .28*** 4) r_p = .22*** 5) r_p = .09 6) r_p = .03 7) r_p = -.21*** 8) r_p = -.29*** 9) r_p = .29*** 10) r_p = .61*** 11) r_p = .50*** 12) r_p = -.54*** 13) r_p = .47*** 14) r_p = -.03	1) r_p = .18*** 2) r_p = .17* 3) r_p = .00 4) r_p = -.10 5) r_p = -.10 6) r_p = .10 7) r_p = -.20*** 8) r_p = .49*** 9) r_p = .34*** 10) r_p = .03 11) r_p = .16*** 12) r_p = .28***

Note. A-DMC = Adult Decision-Making Competence index score (Bruine de Bruin et al., 2007); DAC-test = Decision-Analytical Competence test; DOI = Decision Outcomes Inventory; GPA = grade point average; PA-M = Point average in maths; r_P = Pearson correlation.

* $p < .05$; ** $p < .01$; *** $p < .001$.

The DAC-test is validated by a set of 12 measures: seven self-assessment scales, two performance tests, and three socio-demographic variables. Significant correlations could be observed for decision self-esteem (r_P = .18; p < .001), vigilant decision-making style (r_P = .17; p < .05), and interlerance of uncertainty (r_P = -.20; p < .001). The relationships with fluid intelligence (r_P = .49; p < .001) and problem-solving competence (r_P = .34; p < .001) appeared as the highest values. Also, the overall DAC-test overall score correlates positively with the GPA (r_P = .16; p < .001) and the point average in maths (r_P = .28; p < .001).

It can be summarised that both measures, the A-DMC and the DAC-test, intend to operationalize DMC. The two performance tests differ in terms of their theoretical basis of captured decision-making approach and target group. This section intended to descriptively point out the strengths and weaknesses of the A-DMC and the DAC-test. In direct comparison, the A-DMC seems to perform better in terms of internal consistency on sub-scale and construct-level. According to the construct's dimensionality, the DAC-test shows a better fit of the empirical data in respect to its theory. Validity-wise, the evidence of the A-DMC shows more significant and more distinctive relationships with corresponding validation criteria.

Addressing the strengths and weaknesses of the developed DAC-test, the following two sub-sections discuss the scientific contribution and potential practical utilisation as well as the limitations of the current research and points out potential for further research.

6.4 SCIENTIFIC CONTRIBUTION AND POTENTIAL PRACTICAL UTILISATION

The present research dedicated itself to a relatively under-studied branch of decision-analytical research, which has traditionally emphasised the development of methods and tools to support decision-making. It followed the examples of behavioural approaches, which have been constructed performance-based measures that allow for assessing an individual's decision-making competence on various cognitive dimensions. While these existing instruments (cf. Bruine de Bruin et al., 2007; Parker & Fischhoff, 2005) mostly focus on capturing the ability to resist decision biases and therefore concentrate on roots from descriptive decision science, this PhD research provides a new approach to quantifying individuals' competence to run an analytical decision process to solve complex decision problems. The following paragraphs try to assess the scientific contribution of this study and identify potential areas for its practical utilisation.

Over fifty years since the discipline of decision analysis has emerged, it seemed reasonable to make use of the great amount of knowledge the scientific field has created on how decision makers can better solve complex decision problems. The time appeared to be suitable to focus on individual choices and provide a prescriptive approach to the question of how the quality of individuals' decision-making could be assessed objectively.

By doing a literature review on decision-analytical descriptions of good decision-making processes, eight cognitive steps, which are widely recognised in the field, were identified. The so-called Process Cycle of Decision-Analytical Competence (cf. Figure 2-5) illustrates the aspects of the constructs and thereby defines DAC. The core of the present work was the operationalization of those aspects of DAC. For this purpose, a set of cognitive performance tasks was developed. The thereby designed DAC-test was then field-tested and examined in two studies in terms of its statistically quality criteria.

In the field of decision analysis, this research makes a clear contribution to understanding how researchers of prescriptive decision science define good individual decision-making and how the quality of decision-analytical knowledge can be evaluated by a rigorous assessment tool. So far there are very few attempts of conducting such assessments in the decision analysis literature. One recently published exception is the work of Siebert and Kunz (2016) which proposes the proactive decision-making scale. The scale classifies decision makers' decision-making habits as proactive or reactive by a self-assessment questionnaire. Besides the study of Finucane and Gullion (2010), which provides among other constructs indices of comprehension and dimension weighting, the present research on DAC is the first attempt to offer a performance-based measure to objectively evaluate the quality of individual decision-analytical processes.

The developed DAC-test goes along with a wide range of potential areas for application: Accessible to the public, an individual could use the test results for personal reflection and thereby identify areas for improvement. Corresponding training could be undertaken to address the specific DAC-ability/ies, which showed potential for improvement. Also, the individual would learn more about their strengths and thus, expand their awareness of well-developed abilities.

At the same time, training, workshops, and courses targeting decision-analytical knowledge transfer, might use the DAC-test to understand the status quo of their target group better or run pre-post testing to evaluate their work.

Given the case that various educational programmes used the DAC-test for their evaluation or impact measuring, the effectiveness of different programmes could be compared and thus, clients had the chance to justify their decision for a specific programme on an objective/independent criterion.

One relatively obvious area for application might be the personal selection in human resources. The DAC-test could be used to diagnose differences between people of certain groups. As part of an assessment centre the DAC-test could be applied to figure out who of the applicants showed the highest decision-analytical knowledge. For a job position, which goes along with a high frequency of high-stake decisions, this testing instrument could help to find inter-individual differences and thereby identify the best fitting candidate. Based on the relevance of decision-making competence for the requirements profiles of current and future occupational fields (for instance in the areas of politics, risk management, economics and news selection in fast digital media), the necessity of an instrument to measure this competence is increasing.

In contrast to this section but also reflecting the present research, the next section discusses the limitations and potential for further research.

6.5 LIMITATIONS AND POTENTIAL FOR FURTHER RESEARCH

The present research intended to psychometrically capture DMC from a decision-analytical perspective. Therefore, the latent construct of DAC built the focus of this Ph.D. project. DAC was theoretically defined by the decision-analytical literature and operationalized by a set of performance-based decisions tasks. According to prescriptive decision theory, eight cognitive and motivational dimensions of DAC were revealed: the ability to recognise a decision, the ability to assess decision fitness, the ability to frame a decision, the ability to envision one’s objectives, the ability to identify relevant alternatives, the ability to deal with uncertainty, the ability to integrate information, and the ability to plan to implement a decision. For all DAC-dimensions except the ability to plan to implement a decision a set of objectively, and by a programmed feedback scheme, ad hoc assessable items could be developed.

The remaining seven DAC-dimensions were qualitatively pilot-tested, quantitatively pre-tested, and finally tested in the main study. Due to weakly observed variance within the data set for the ability to envision one’s objectives, no empirical evidence for this sub-scale of DAC could be collected. The other dimensions were examined in terms of their internal consistency and dimensionality. On a construct-level, DAC was analysed in terms of its internal consistency, dimensionality, and validity. The latter quality criterion was investigated by a set of 12 different validation criteria.

The following sub-section serves to highlight subset aspects and results of the current research in terms of their limitations and potential for further research. Sub-section 6.5.1 discusses aspects of the data quality and transferability of the research results. Sub-section 6.5.2 reflects the statistical properties of the developed DAC-test and the potential for improvement. Sub-section 6.5.3 points out that not all intended dimensions of DAC could be empirically captured by this Ph.D. research and sub-section 6.5.4 presents prospects for future research on measuring DAC.

6.5.1 Aspects of the Data Quality & Transferability of Results

The main reason for designing an online test rather than a paper-pencil test was the limited resources of a Ph.D. research – time- and money-wise. Thus, the developed DAC-test set up as an online measure helped to keep the expenses for the acquisition of test participants, for printing, and for data entry low. It provided great speed in the phases of acquisition, responding, and data entry. Also, a positive side-effect was the corresponding objectivity (cf. sub-section 5.2.5) of data entry and evaluation.

Even though the challenges of online studies, such as multiple submissions, incomplete responses, or high dropout rates (Reips, 2001), have be addressed in preparation of the research, a general sample bias could not be prevented (Tuten, Urban, & Bosnjak, 2002). For the pre-testing, approximately 2190 different users entered the DAC-test website, of whom 8.95 per cent (N = 196) completed the test. For the main testing, approximately 3430 various IP-addresses accessed the website within the period of data gathering. Of those web-site visitors, 13.27 per cent (N = 433) completed the test and thereby submitted their results. To test whether the sample group of the present research shows obviously group differences, maybe due to sample biases, four group comparison tests have been undertaken. With regard to the *classic* socio-demographic variables, sex, age, and education, a quite homogeneous picture appears. An independent samples t-test was conducted to compare the scaled overall DAC-test scores for male and female test participants. No significant difference between female test participants (M = 0.11, SD = 0.95) and male test participants (M = -0.02, SD = 1.03) was found; t (353) = 1.20, p = .231. By a Pearson correlation, the relation among age and the scaled overall DAC-test score was examined. No significant correlation was found with r_P = .071 (p = .182). A nonparametric correlation with Spearman's rho was calculated to analyse the coherence of the scaled overall DAC-test score and the educational level of the test participants. Also here, no significant difference was found by r_S = .061 (p = .252).

Additionally, to investigate whether the two groups of paid and unpaid test participants differ in terms of their performance in the DAC-test, again an independent samples *t*-test was run. The groups did not differ significantly. Paid participants ended up with a scaled overall mean DAC-test score of M = 0.02 (SD = 1.02) on average and unpaid test participants reached a mean overall DAC-test score of M = 0.08 (SD = 0.95); t (355) = 0.63, p = .530.

A general issue that had to be addressed for the present research is its quality of data. Nevertheless, given the advantages of online testing for the researcher, e.g. efficiency and access to a large locally-independent sample (e.g. Dillman et al., 2014), and for test participants, e.g. their comfort to complete the test whenever and wherever they want, less control over the quality of data might be a price to pay. One way to control for data quality was the set control criterion of a minimum completion time for the survey as a whole. However, the criterion of minimum completion time turned out to be not distinctively applicable as 7.24 per cent of test participants, who attained more than 90 per cent of the overall DAC-test score, fell below the set threshold of 30 minutes for completion. Other attempts at data cleansing, such as by controlling for motivation or level of English skills, did not bring any clearer picture of the empirical data. To better address the issue of data quality in further research on this topic, standardised and controlled testing in labs might be an opportunity.

The results of the present research have to be considered more as a work in progress than as final statements about how DAC has to be measured for young adults. This Ph.D. research is mainly an attempt to create a psychometric performance measure and especially in terms of its statistical properties it became clear that further research effort is needed. The current developed DAC-test is tailored towards young people between 18 and 30 years. The test is presented in English, and thus demands a proper level of English understanding. That is why people with self-assured “basic English skills” were excluded from the data analyses. The results are exclusively assigned to the observed sample group and consequentially cannot be used to draw a conclusion for the population in general.

6.5.2 Improving Statistical Properties for the DAC-Test

As consistently pointed out during this Ph.D. research, its main aim was to create a reliable and valid DAC performance test. On the basis of a literature review, analytical steps for a sound decision-making process were identified and operationalized by a set of either existing, adapted, or newly created cognitive decision tasks. At the very beginning the empirical research intended to prove the reliability of the designed test by examining two parallel-tests of the DAC-test. Therefore, two comparable tests were developed, pilot-, and pre-tested. However, the actual completion time for both test forms and the corresponding effort for test participants exceeded the level of reasonableness with over three hours on average. That is why one test form had to be neglected and only one of the two test forms could be provided and tested in the main testing. Consequently, testing for parallel-test reliability was not possible. The measure of split-half reliability was indirectly determined by the mean of all split-half measures expressed by Cronbach's alpha. This coefficient in turn only works for homogeneous scales and is very susceptible for scales with fewer than 10 items (Nunnally et al., 1978), as occurred for five of the six tested DAC-test sub-scales. Therefore, it might be advisable to provide for each DAC-test sub-scale at least 10 items; ideally, the same number of items so that a comparison of the DAC sub-scales' Cronbach's alpha provides more effective information on the internal consistency and does not leave so much room for conjecture. Further research is therefore necessary to expand the examination the reliability of the DAC-test and its sub-scales and thereby improve the statistical properties. Especially, gathering evidence on the DAC-test's re-test reliability would be worthwhile.

Additionally, to further improve the comparability of the DAC-dimensions, it might be reasonable to provide all items of the DAC-test with the same number of responses. In this case, the chances of *guessing correctly* would be comparatively equal over all items and thus, could be left out from analysing considerations.

In terms of the dimensionality on a construct level, satisfying results have been extracted by an EFA and confirmed by a CFA. Hence, as assumed in relation to the literature (Stanovich & West, 2000), DAC can be considered as a construct underlying several correlated dimensions. The data of the present research provided undetermined results on the homogeneity of the DAC-test sub-scales. Therefore, further exploration on the dimensionality of the DAC-dimensions is necessary[81].

Referring to the test's validity, significant correlations in expected direction to the theory could be observed for the seven constructs and variables: decision self-esteem, vigilant decision-making style, intolerance of uncertainty, fluid intelligence, problem-solving competence, the GPA, and the point average in maths. These results are treated as first evidence for the validity of the developed DAC-test. Nevertheless, further validation is

---

[81] Besides some possible irregularities in the item answers due to "clicking-through", the CFA results point to another topic for re-evaluation. To increase the statistical properties of the scales it might be necessary to address the issue that the items of some scales show low discriminatory power or plain lack of unidimensionality and to further scrutinise these items. This may lead to another cycle of item construction and evaluation in order to replace some of the existing wordings. Despite the effort undertaken in this thesis to validate the DAC-test, this task is not completed yet and further research is required from a statistical point of view.

essential. So, an explicit and more distinct ascription of DAC to sound decision-making styles and habits is desirable. At the same time a precise differentiation from irrational, emotional, or non-analytical approaches to decision-making is neccessary. In addition, the strong relation to intelligence is not suprising. However, a definite borderline to the construct of intelligence is crucial and hence, calls for further examination.

6.5.3 Providing a More Holistic Picture of DAC

According to prescriptive decision theory, DAC is seen as a multi-dimensional higher-order construct (cf. Finucane & Gullion, 2010; Stanovich & West, 2000). The decision-analytical literature captures a set of cognitive steps that describes a good decision-making process (cf. Mellers & Locke, 2007). The DAC-test aimed to make all relevant dimensions measureable by an online performance test. However, some constraints mainly in terms of data analysis were the reason for why not all eight identified DAC-dimensions could be successfully operationalized and measured.

The following remarks address the missing aspects of DAC and present ideas on how a more holistic picture of the construct could be operationalized in future research.

Ability to Envision One's Objectives

One aspect of analytical decision-making, which is indispensable to decision analysis, is the ability to envision one's objectives - OBJECTIVES. However, both in the pre-testing and in the main testing, the data of OBJECTIVES had to be excluded from further analyses as for both test score distributions a ceiling effect appeared (Cramer & Howitt, 2004), with which the condition of a normal distribution was violated and variances within the group could not be observed sufficiently. After adapting the number of required objectives from seven in the pre-testing to 10 in the main testing, it was assumed to be able to correct this effect, as study participants of Bond et al. (2008) have on average selected 7.7 self-generated objectives of the final chosen objectives (cf. Table 6-6). By increasing the required number to 10 objectives, it was intended to create more room for variance of the test results. Table 6-6 gives an overview of OBJECTIVES' indicators of the present research and the study of Bond et al. (2008).

Table 6-6 Comparing Bond et al. (2008) and the DAC-test results for OBJECTIVES *indicators*

	Bond et al., 2008				Present study		
	Study 1	Study 2	Study 3	AV	Pre-testing	Main testing	AV
Mean number of listed objectives	7.4	5.9	6.8	**6.7**	8.0	9.7	**8.9**
Mean number of objectives checked as relevant	13.6	14.3	14.8	**14.2**	19.9	22.9	**21.4**
Mean number of self-generated objectives of the final selected objectives (percentage)	7.6 (76%)	7.7 (77%)	7.9 (77%)	**7.7 (77%)**	5.5 (79%)	8.5 (85%)	**-[a] (82%)**

Note. DAC = Decision-analytical competence; AV = average.

[a] The dash indicates that an average of the means of self-generated objectives of the final selected objectives of pre-testing and main testing is not applicable as the number of required objectives had been increased from seven to 10 and thus might have anchored test participants. In the Bond et al. studies each time the selection of 10 objectives was required.

It becomes clear that the results of both research projects do not differ that much, even though the percentage of self-generated objectives of the final selected objectives (cf. last row) is five per cent higher in the current study.

The crucial difference between the two studies is their corresponding research intentions. While Bond and colleagues aimed to prove that individuals are not able to envision a complete list of individually relevant objectives, the present study wanted to measure inter-individual differences. Due to the received ceiling-effect and thereby the missing variances within the group, neither the setting with seven nor the setting with 10 required objectives fitted this research aim. Deductively, the operationalization of OBJECTIVES as it was applied for this research seems to be too *easy* for this group of test participants and for this reason another way of operationalization has to be found in further research.

(Nisbett and Wilson (1977) analysed introspection in think-aloud experiments and came to the conclusion that people are not able to achieve true introspection when asked to report on their cognitive processes. Thus, OBJECTIVES might be a challenging task per se for test takers and might not properly measure the intended construct rather introspection. Nevertheless, the approach of Bond, Carlson, & Keeney (2008) was identified as the best trial so far to operationalize this important aspect of DAC.

One further critical aspect to the operationalization of this DAC-dimension is its assessment criterion. In comparison to the assessment criteria of the other DAC sub-scales (cf. Table 6-1), for which the accuracy of test participants' responses could be evaluated objectively, for this decision task the percentage of self-generated objectives of the finally selected, individually most important, objectives is calculated. Some test participants might have identified this assessment criterion. One possible way to reduce this potential effect in further studies, is to divide the OBJECTIVES task: to ask at the beginning to list all objectives that arise in the test participant's mind, to choose in the second step the relevant objectives from the master list, and to match the own list with the master list. After completing the remaining tasks of the DAC-test, i.e. at the very end of all items, test participants are required to choose their 10 most important objectives from the combined list of self-generated and identified objectives. To support this setting, it should not be traceable for test participants which of the objectives of the final list have been identified as self-generated in the first step and which have not been identified.

Ability to Plan to Implement a Decision

As presented in sub-section 4.1.8, the first ideas of how the ability to plan to implement a decision - IMPLEMENTATION - could be operationalized were collected. However, within the framework of this Ph.D. no standard of comparison for the qualitative results of the pre-testing for this DAC-test sub-scale could be extracted. For the sake of completeness, future research should address this issue and find a way of quantitatively assessing the ability to plan to implement a decision.

Even though the other six DAC-dimensions effectively have been operationalized quantitatively and the corresponding results turned out analysable, enhancements for two of the sub-scales are presented in the following, as additional aspects of these two dimensions should be considered in further research of DAC.

Ability to Assess Decision Fitness

Referring to decision-analytical literature, the ability to assess decision fitness – FITNESS - implies a skill and can be described as metacognition (W. Edwards, 1954; Raiffa, 1968), i.e. being aware of the extent of one's competences. So far, this aspect has not been considered for measuring DAC. For prospective studies, one indicator could be added to the DAC-test. This indicator would measure whether people are able to realistically assess the extent of their cognitive competences after having solved a decision-making task. Such an item would be a self-evaluation task. In this respect, it could be set right after each item block of each DAC-dimension. Test participants would be asked to estimate their performance on the completed block of items. This type of item could serve to measure whether test participants are able to assess the quality of their own decision-making competency and thereby provide a more holistic picture of this DAC-dimension.

Ability to Identify Relevant Alternatives

The ability to identify relevant alternatives - ALTERNATIVES - is hitherto measured by items that request test participants to decide whether provided alternatives fit presented individual objectives of the character in the test. This type of item refers to the qualitative aspect of this DAC-dimension, as test participants have to evaluate the alternatives on the basis of the given objectives. An additional aspect of ALTERNATIVES could be quantity. According to the literature, good decision-making requires the creation of a sound set of alternatives (Siebert & Keeney, 2013). As a first step, test participants could be asked to list as many alternatives as they can think of for a presented decision situation, which is connected to the main case of the test. This type of item would refer to the quantitative aspect of this DAC-dimension and could be split into creativity, which is defined by originality, and fluency, which describes "the ability to generate ideas, words, mental associations, or potential solutions to a problem with ease and rapidity." (APA, 2007, p. 381 [entry "fluency"])

6.5.4 Future Prospects for Researching DAC

Summarising, the results of the present Ph.D. research show promise for capturing analytical DMC of individuals as a construct underlying related cognitive dimensions. The empirical findings verify this assumption. However, further research intensions are necessary to further develop the assessment to a more advanced level. Some future attention should be paid to analyse the dimensionality of DAC and the structure of the DAC-dimensions. In this respect, effort should be undertaken to expand the number of solid items with comparable response sets. Especially in terms of statistical properties, the DAC-test needs additional mindfulness. So, the reliability of each dimension and the DAC-test as a whole has to be proven, also by re-test reliability. More distinct and rigorous validation criteria have to be applied. A distinction among emotional, intuitive, and irrational decision-making approaches and intelligence is essential. Former studies have found positive correlations between DMC related measures and tests for aspects of intelligence (e.g. Bruine de Bruin et al., 2007). In this context, a positive correlation of DAC and IQ would indicate convergent validity. However, further studies have to prove the distinction between DAC and fluid intelligence so that it becomes explicit that the DAC-test provides additional benefit in comparison to common tests for fluid intelligence.

DAC could be seen as one aspect of DMC or the combination of various analytical abilities of the general construct of DMC. This relation could be seen as one reason why the correlation of DAC and the scales of the MDMQ-II (Mann et al., 1997) are not so high. Direction for further research should include more decision-analytical specific criteria for validation. One potential scale developed recently[82] is the proactive decision-making scale of Siebert and Kunz (2016). However, as DAC intends to capture parts of the decision-analytical components of DMC a correlation of DAC and DMC and correspondingly DMC-related constructs was assumed.

To provide a more holistic picture of the construct, OBJECTIVES has to be tested in a way that the test results show enough variance that could be analysed. Also, a method has to be found to quantitatively operationalize IMPLEMENTATION and thereby complement the DAC-test by capturing all eight cognitive dimensions. Even though the present research presents impetuses to assess FITNESS and ALTERNATIVES quantitatively, future studies could be dedicated to the research of different aspects of these two DAC-dimensions.

Having overcome these obstacles, it would definitely be worthwhile to examine the already designed parallel-test, which had to be rejected after the pre-testing due to time constraints.

Furthermore, a joined study measuring A-DMC, which captures the behavioural decision science side, and the DAC-test, which measures the prescriptive side of decision-making, would be of interest, as a good decision-making process involves the ability to deal with decision biases and the ability to analytically solve a decision problem (e.g. Baron, 2008).

[82] The proactive decision-making scale was published 2016. Unfortunately, this was too late to include it in the present research and thereby use it as a criterion for validation.

7 CONCLUSION

Decision-making builds key competences - in terms of business (e.g. Hoffman et al., 2011) as well as in private lives (e.g. Keeney, 2008). Various research groups commit themselves to the interdisciplinary topic of decision-science. Within this context, the descriptive approach with its roots in behavioural psychology investigates how people naturally solve decision problems (e.g. Gilovich, Griffin & Kahneman, 2002; Takemura, 2014); the normative approach originates from decision theory and analyses how choices can be made ideally (e.g. Edwards, Miles Jr., & von Winterfeldt, 2007), and the prescriptive approach connects the two prior research fields and is dedicated to the analysis of decision-making processes, helping people to improve their decision-making by developing efficient decision supporting systems and strategies (e.g. D. E. Bell et al., 1988; Keeney, 1992).

Due to this diversity of research interests, a variety of approaches for defining decision-making competence (DMC; Finucane & Lees, 2005) have been proposed in the literature. Nevertheless, no consistent definition of DMC with a stated set of related dimensions is generally accepted, even though decision scientists, such as Milkman et al. (2009) or Stanovich and West (2000), assume that DMC builds a higher-order factor underlying a set of cognitive abilities.

Psychometric instruments to measure DMC and especially performance-based measures are still rare. Concomitantly, attempts to operationalize DMC from the decision-analytical avenue of decision theory in a theory-driven manner were missing.

The present dissertation aimed to address this gap of a lack of a performance test to measure the quality of individual decision-analytical competence (DAC). Hence, the main goal of this research was to create a theory-driven and evidence-based test with good statistical properties quantifying DAC on an individual level.

The thesis had four main research objectives. The first research objective was to identify the strengths and weaknesses of prevalent definitions and measures of DMC and identify gaps for improvement. Besides a variety of attempts on measuring DMC by self-rating scales (e.g. Mann et al., 1997; Mann, Harmoni, Power, et al., 1988), most recent research operationalizes DMC by a set of performance-based tasks. The Adult Decision-Making Competence index score (A-DMC) of Bruine de Bruin et al. (2007), which focuses on findings from behavioural psychology and normative decision science, appears as a promising approach to the perspective of DMC as a higher-order construct. Since other performance tools, such as the work of Finucane and Gullion (2010), are often situation-specific, the A-DMC was identified as one of the most advanced approaches in this field trying to capture general DMC. The perhaps greatest potential for improvement for developing a new test instrument to measure DMC lies in the fact that the A-DMC only focuses on the assessment of abilities to deal with decision biases and thus, misses aspects derived from prescriptive decision science.

For this reason, the second research objective was the conceptualisation of DAC as a psychological construct, consisting of a concrete number of measurable dimensions according to the theory of decision analysis. Introduced in this context was the *Process Cycle of Decision-Analytical Competence* with its eight dimensions, deduced from decision-analytical literature (e.g. Bazerman & Moore, 2009; W. Edwards, 1954; Hammond et al., 1999; Hastie & Dawes, 2010; Howard, 2007; Keelin et al., 2009; Keeney, 1992; Siebert & Keeney, 2015). These eight dimensions embrace the ability to recognise decision opportunities (OPPORTUNITIES), the ability to assess decision fitness (FITNESS), the ability to frame a decision (FRAME), the ability to envision one's objectives (OBJECTIVES), the ability to identify relevant alternatives (ALTERNATIVES), the ability to deal with uncertainty (UNCERTAINTY), the ability to integrate information (INFORMATION), and the ability to plan to implement a decision (IMPLEMENTATION), and eventually build the theoretical basis for the DAC-test construction.

The development of corresponding item sets to operationalize the various theoretical DAC-dimensions and in parallel, the construction of a psychometric performance test to measure DAC, formed the third main research objective. Test constructional considerations led to the decision of presenting the DAC-test as an online test, which consists of content-wise connected items, in a wider sense capturing the complex decision situation everyone faces at least once in their lifetime "What to do after having completed school". Young adults between 18 and 30 years built the target group for the DAC-test study.

For all DAC-dimensions, except IMPLEMENTATION, a set of quantitative assessable items was developed and pre-tested on a sample group of 143 young adults. The pre-testing served to examine the quality of items and the dimensionality of the construct. The corresponding exploratory factor analysis with non-orthogonal promax rotation extracted a one-factor solution explaining 29% of the observed variance, which supported the theoretical assumption of DAC being a higher-level measure, underlying a set of related dimensions.

Subsequently the fourth research objective was to examine the proposed psychometric DAC-test in terms of its reliability and validity relating to its empirical evidence for sound decision-making as measured by a set of appropriate criteria for validation. In the main testing, the final version of the DAC-test was examined by 368 test participants, of which 47.5% were male and 52.7% female. On average, test participants were 24 years old (*M* = 23.8; *SD* = 3.4). The 49 items of the six dimensions, which showed sufficient variance in the data[83], presented a Cronbach's alpha of α = .85 and the z-scores of the sub-scales sum-scores a McDonald's omega of ω_H = .72 for internal consistency. While the scales of FITNESS, FRAME, and ALTERNATIVES were homogeneous, the other DAC-test sub-scales were not. The DAC-test sub-scales of OPPORTUNITIES, UNCERTAINTY, and INFORMATION showed internal consistency of $\omega_H > .70$.

The assumed one-factor structure for DAC was confirmed by a confirmatory factor analysis showing a good model fit (CFI = .974, RMSEA = .051, SRMR = .031). Thus, according to the theory, the results of this research reinforced DAC, measured by the DAC-test, as a higher-level construct underlying a set of correlated dimensions.

The validation of the DAC-test was examined by a set of 12 measures: seven self-assessment scales, two performance tests, and three socio-demographic variables. Significant correlations could be observed for decision self-esteem (r_P = .18; $p < .001$), vigilant decision-making style (r_P = .17; $p < .05$), and interlorance of uncertainty (r_P = -.20; $p < .001$). The relationships with fluid intelligence (r_P = .49; $p < .001$) and problem-solving competence (r_P = .34; $p < .001$) appeared highest. Also, the DAC-test overall scores correlated positively with the GPA (r_P = .16; $p < .001$) and the point average in maths (r_P = .28; $p < .001$). Thus, the results are treated as first evidence for the DAC-test's validity.

However, further research is necessary to examine the reliability of the DAC-test, its sub-scales, and in parallel, the sub-scales' dimensionality, as the results of this study are not distinct for all corresponding coefficients. It would also be worthwhile to gather evidence on the DAC-test's re-test reliability, as the present research did not cover this quality criterion. Additionally, further research should extend the catalogue of convergent and discriminant validation criteria for the DAC-test, and thereby elaborate a clear attribution of the DAC-test to approaches of analytical decision-making and provide a precise distinction from non-analytical approaches. The high correlation of DAC and fluid

[83] Due to a ceiling effect and thus missing variance within the test data, OBJECTIVES had to be excluded from the final analyses.

intelligence also needs more attention, resulting in a crucial contour between the two concepts.

Referring to the main aim of the study, the present research should be treated as a first step in the direction of performance measures for DAC concerning complex decisions for the individual. It is desirable that further research attempts are undertaken to expand this work. A widely accepted, valid, and reliable psychometric performance test capturing the quality of DAC would build the basis for three application areas and thereby show high relevance for practitioners. Firstly, for the individual per se, it would allow for measuring and evaluating DAC on different dimensions. In this respect, it would be possible to provide detailed individual feedback about how well each ability of DAC is evolved. Courses and workshops could then respond to this assessment and address the gaps of the decision-maker. Secondly, for inter-individual differences, a comparison of individual performances in the processing of the DAC-test would allow for differentiating between decision-makers according to their performance quality of DAC. Consequently, the DAC-test could be applied within assessment centres by human resource departments to examine candidates' level of DAC and thereby their qualification for certain occupational requirements. Thirdly, for measuring change, a DAC-test would allow for testing DAC before and after training units. Therefore, teachers, lecturers, or workshop leaders could gain data on students' progress and so evaluate the impact of their work.

Summarising, the present work is an attempt to measure DAC in a rigorous and systematic way. The constructed DAC-test shows promising first evidence of reliability and validity. Gaining more empirical insights on how to measure and assess DAC and thereby move this research topic further towards researchers and practitioners would help to establish DAC as a psychological construct. Further research to improve the DAC-test is of great importance as providing a performance test with good statistical properties could arouse increasing interest in domains as diverse as science, economics, and politics.

Undertaking this journey is worth it, given the importance of decisions to individual and societal success and wellbeing, because "[t]he only way to exert control over your life is through your decision making. The rest just happens to you." (Hammond et al., 1999, p. 234)

8 BIBLIOGRAPHY

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). New Jersey: Prentice-Hall, Inc.

Anderson, J. R., & Douglass, S. (2001). Tower of Hanoi: Evidence for the cost of goal retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1331–1346. http://doi.org/10.1037/0278-7393.27.6.1331

Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, *6*(3), 252–262. Retrieved from http://journal.sjdm.org/11/11218/jdm11218.html

ARD Forschungsdienst. (2014). Mediennutzung von Jugendlichen [Media use of the youth], 242–248.

Bandura, A., & Walters, R. H. (1963). *Social learning and personality development*. New York: Holt Rinehart and Winston.

Baron, J. (2008). *Thinking and deciding* (4th ed.). New York, NY: Cambridge University Press.

Baron, J., & Brown, R. V. (1991). *Teaching decision making to adolescents*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Barron, G., & Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, *18*(5), 507–520. http://doi.org/10.1016/S0747-5632(02)00007-9

Bazerman, M. H., & Moore, D. A. (2009). *Judgment in Managerial Decision Making* (7th ed.). Hoboken, HJ: John Wiley & Sons, Ltd.

Bell, D. E., Raiffa, H., & Tversky, A. (1988). Descriptive, normative, and descriptive interactions in decision making. In *Decision making - descriptive, normative, and prescriptive interactions*. Cambridge: Cambridge University Press.

Bell, H. J. (2004). Respiratory response to passive limb movement is suppressed by a cognitive task. *Journal of Applied Physiology*, *97*(6), 2112–2120. http://doi.org/10.1152/japplphysiol.00302.2004

Berkemeyer, N., Bos, W., Manitius, V., Hermstein, B., Bonitz, M., & Semper, I. (2014). *Chancenspiegel 2014: Regional Disparitäten in der Chancengerechtigkeit und Leistungsfähigkeit der deutschen Schulsysteme [Regional disparities of fairness of opportunites and achievement potential of Germany's school systems]*. Gütersloh.

Bertsch, M., Huth, N., & Arenz, R. (2011). *Netzgesellschaft: Eine repräsentative Untersuchung zur Mediennutzung und dem Informationsverhalten der Gesellschaft in Deutschland [A study on media use and information patterns in Germany]*. Berlin.

Birnholtz, J. P., Horn, D. B., Finholt, T. A., & Bae, S. J. (2004). The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents. *Social Science Computer Review*, *22*(3), 355–362. http://doi.org/10.1177/0894439304263147

Birrell, J., Meares, K., Wilkinson, A., & Freeston, M. (2011). Toward a definition of intolerance of uncertainty: A review of factor analytical studies of the Intolerance of Uncertainty Scale. *Clinical Psychology Review*, *31*(7), 1198–1208. http://doi.org/10.1016/j.cpr.2011.07.009

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, *314*(7080), 572–572. http://doi.org/10.1136/bmj.314.7080.572

Boleman, L. G., & Deal, T. E. (1991). *Reframing organizations: Artistry, choice, and leader* (4th ed.). San Francisco, CA, USA: Jossey-Bass Publishers.

Bond, S. D., Carlson, K. a., & Keeney, R. L. (2008). Generating objectives: Can decision makers articulate what they want? *Management Science*, *54*(1), 56–70. http://doi.org/10.1287/mnsc.1070.0754

Bond, S. D., Carlson, K. a., & Keeney, R. L. (2010). Improving the generation of decision objectives. *Decision Analysis*, *7*(3), 238–255. http://doi.org/10.1287/deca.1100.0172

Bortz, J., & Döring, N. (2005). *Forschungsmethoden und Evaluation: für Human- und*

Sozialwissenschaftler [Research methods and evaluation for human and social sciences] (3rd ed.). Berlin: Springer.

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in web surveys: an experiment. *Social Science Computer Review*, *21*(2), 208–217. http://doi.org/10.1177/0894439303021002006

Bransford, J., & Stein, B. (1984). *The IDEAL problem solver*. New York, NY: W. H. Freeman.

Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structure equation models* (pp. 62–136). Newbury Park, CA: SAGE Publications.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*(5), 938–56. http://doi.org/10.1037/0022-3514.92.5.938

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2012). Explaining adult age differences in decision-making competence. *Journal of Behavioral Decision Making*, *25*(4), 352–360. http://doi.org/10.1002/bdm.712

Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion [Introduction to test and questionnaire construction]* (3rd ed.). München: Pearson.

Buhr, K., & Dugas, M. . (2002). The intolerance of uncertainty scale: Psychometric properties of the English version. *Behaviour Research and Therapy*, *40*(8), 931–945. http://doi.org/10.1016/S0005-7967(01)00092-4

Burnett, P. C., Mann, L., & Beswick, G. (1989). Validation of the flinders decision making questionnaire on course decision making by students. *Australian Psychologist*, *24*(2), 285–292. http://doi.org/10.1080/00050068908259567

Butler, A. B., & Scherer, L. L. (1997). The effects of elicitation aids, knowledge, and problem content on option quantity and quality. *Organizational Behavior and Human Decision Processes*, *72*(2), 184–202. http://doi.org/10.1006/obhd.1997.2737

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. http://doi.org/10.1037/h0046016

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. http://doi.org/10.1037/0033-295X.97.3.404

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. http://doi.org/10.1037/h0046743

Cattell, R. B. (1965). *The Scientific Analysis of Personality*. London, UK: Penguin Books.

Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action*. (G. E. Stelmach & P. A. Vroon, Eds.). North-Holland: Elsevier Science Publishers B.V.

Chaiken, S., & Trope, Y. (1999). *Dual process theories in social science*. New York: Guilford Press.

Cheek, J. M., & Hogan, R. (1983). Self-concepts, self-presentations, and moral judgments. *Psychological Perspectives on the Self*, *2*, 249–273.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Oxford, UK: Prentice-Hall, Inc.

Couper, M. P. (2008). *Designing effective web surveys*. Cambridge: Cambridge University Press.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*(2), 230–253. http://doi.org/10.1086/322199

Cramer, D., & Howitt, D. L. (2004). *The SAGE dictionary of statistics: A practical resource for students in the social sciences* (3rd ed.). London, UK: SAGE, Publications, Ltd.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Orlando, FL: Wadsworth Pub Co.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. http://doi.org/10.1007/BF02310555

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. http://doi.org/10.1037/h0040957

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, *40*(2), 428–434. http://doi.org/10.3758/BRM.40.2.428

Dawes, R. M. (1988). *Rational choice in an uncertain world*. New York, NY: Harcourt Brace Jovanovich.

Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking & Reasoning*, *16*(2), 69–97. http://doi.org/10.1080/13546781003630117

Deniz, M. E. (2006). The relationships among coping with stress, life satisfaction, decision-making styles and decision self-esteem: An investigation with turkish university students. *Social Behavior and Personality*, *34*(9), 1161–1170. http://doi.org/10.2224/sbp.2006.34.9.1161

Dieckmann, N. (2008). Numeracy: A review of the literature. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.1561876

Diener, E., Emmons, R. a, Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–5. http://doi.org/10.1207/s15327752jpa4901_13

Dillman, D. A., Smyth, J. D., & Melani Christian, L. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). New Jersey: John Wiley & Sons, Ltd.

Dugas, M. J., & Robichaud, M. (2012). *Cognitive-behavioral treatment for generalized anxiety disorder: From science to practice*. New York, NY: Taylor & Francis Group.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., … Wechsler, H. (2010). Youth risk behavior surveillance - United States, 2009. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *59*(5), 1–142. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20520591

Edwards, A. L. (1982). *The social desirability variable in personality assessment and research* (reprint). Worth, TX: Greenwood Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. http://doi.org/10.1037/1082-989X.5.2.155

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417. Retrieved from http://psycnet.apa.org/journals/bul/51/4/380/

Edwards, W. (1984). What constitutes “a good decision”? *Acta Psychologica*, *56*, 5–27.

Edwards, W., Miles Jr., R. F., & von Winterfeldt, D. (2007). Introduction. In W. Edwards, R. F. Miles Jr., & D. von Winterfeldt (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 1–13). New York, NY: Cambridge University Press.

Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion [Test theory and test construction]*. (E. Bamberg, H.-W. Bierhoff, A. Grob, & F. Petermann, Eds.). Göttingen: Hogrefe Verlag GmbH & Co. KG.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, *73*(3), 221–230. http://doi.org/10.1037/h0028695

Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin*, *43*(5), 385–440. http://doi.org/10.1037/h0055483

Erikson, E. H. (1959). *Identity and the life cycle*. New York, NY: International Universities Press.

Everitt, B. S. (2006). *The Cambridge dictionary of statistics* (3rd ed.). Cambridge, UK: Cambridge University Press.

Fenton, N. E., & Whitty, R. (1995). Introduction. In N. E. Fenton, R. Whitty, & Y. Lizuka (Eds.), *Software quality assurance and measurement - a worldwide perspective* (pp. 1–19). London: International Thomson Computer Press.

Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging*, *25*(2), 271–288. http://doi.org/10.1037/a0019106

Finucane, M. L., Mertz, C. K., Slovic, P., & Schmidt, E. S. (2005). Task complexity and older adults' decision-making competence. *Psychology and Aging*, *20*(1), 71–84. http://doi.org/10.1037/0882-7974.20.1.71

Finucane, M. L., Slovic, P., Hibbard, J. H., Peters, E., Mertz, C. K., & MacGregor, D. G. (2002). Aging and decision-making competence: An analysis of comprehension and consistency skills in older versus younger adults considering health-plan options. *Journal of Behavioral Decision Making*, *15*(2), 141–164. http://doi.org/10.1002/bdm.407

Fischhoff, B. (2008). Assessing adolescent decision-making competence. *Developmental Review*, *28*(1), 12–28. http://doi.org/10.1016/j.dr.2007.08.001

Fisseni, H. J. (2004). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention [Textbook of psychological diagnostics]* (3rd ed.). Göttingen: Hogrefe Verlag GmbH & Co. KG.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford, UK: Brooks/Cole.

Forstmeier, S., Uhlendorff, H., & Maercker, A. (2005). Diagnostik von Ressourcen im Alter [Clinical diagnostics of resources maturity]. *Zeitschrift Für Gerontopsychologie & -Psychiatrie*, *18*(4), 227–257. http://doi.org/10.1024/1011-6877.18.4.227

Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 4–23). Hillsdale, NJ: Erlbaum.

Frisbie, D. A. (1981). The relative difficulty ratio - a test and item index. *Educational and Psychological Measurement*, *41*(2), 333–339. http://doi.org/10.1177/001316448104100210

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241. http://doi.org/10.1111/j.1467-9280.1996.tb00366.x

Furby, L., & Beyth-Marom, R. (1992). Risk taking in adolescence: A decision-making perspective. *Developmental Review*, *12*(1), 1–44. http://doi.org/10.1016/0273-2297(92)90002-J

George, D., & Mallery, P. (2011). *IBM SPSS statistics 19 step by step: A simple guide and reference* (12th ed.). Toronto, Ontario, Canada: Pearson Prentice-Hall.

Gettys, C. F., Pliske, R. M., Manning, C., & Casey, J. T. (1987). An evaluation of human act generation performance. *Organizational Behavior and Human Decision Processes*, *39*(1), 23–51. http://doi.org/10.1016/0749-5978(87)90044-6

Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Medical Decision Making*, *16*(3), 273–280. http://doi.org/10.1177/0272989X9601600312

Gigerenzer, G. (2003). *Calculated risks: How to know when numbers deceive you*. Simon & Schuster.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. http://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623–629. http://doi.org/10.1111/j.1539-6924.2005.00608.x

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilovich, T., Griffin, D, & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. (D. Gilovich, T., Griffin, D, & Kahneman, Ed.). Cambridge: Cambridge University Press.

Goffman, E. (1956). Embarrassment and social organization. *American Journal of Sociology*, *62*(3), 264–271. Retrieved from http://www.jstor.org/stable/2772920

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493–503. http://doi.org/10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology*, *73*(1), 186–199. http://doi.org/10.1037/0022-3514.73.1.186

Gollwitzer, P. M., & Oettingen, G. (1998). The emergence and implementation of health goals.

Psychology & Health, *13*(4), 687–715. http://doi.org/10.1080/08870449808407424

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes (pp. 69–119). http://doi.org/10.1016/S0065-2601(06)38002-1

Goodwin, P., & Wright, G. (2014). *Decision analysis for management judgment*. John Wiley & Sons.

Gorusch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Grisso, T., & Tomkins, A. J. (1996). Communicating violence risk assessments. *American Psychologist*, *51*(9), 928–930. http://doi.org/10.1037/0003-066X.51.9.928

Guadagnoli, E., & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275. http://doi.org/10.1037/0033-2909.103.2.265

Guilford, J. P. (1956). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill.

Guliksen, H. (1950). *Theory of mental test*. New York, NY: John Wiley & Sons, Ltd. Retrieved from https://books.google.de/books?hl=de&lr=&id=HAu-B6vgd38C&oi=fnd&pg=PP2&dq=Statistical+Theories+of+Mental+Test+Scores&ots=UxdxXAJICE&sig=6FF-d80qy7RlqazKZnslKD5_ygg#v=onepage&q=Statistical Theories of Mental Test Scores&f=false

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998). *Multivariate data analysis* (5th ed.). London: Prentice-Hall, Inc.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1999). *Smart choices: A practical guide to making better decisions*. Boston, MA: Harvard Business School Press.

Hanoch, Y., Wood, S., & Rice, T. (2007). Bounded rationality, emotions and older adult decision making: not so fast and yet so frugal. *Human Development*, *50*(6), 333–358. http://doi.org/10.1159/000109835

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, *52*(1), 653–683. http://doi.org/10.1146/annurev.psych.52.1.653

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making* (2nd ed.). Thousand Oaks, CA: SAGE, Publications, Inc.

Hayashi, M., Kato, M., Igarashi, K., & Kashima, H. (2008). Superior fluid intelligence in children with Asperger's disorder. *Brain and Cognition*, *66*(3), 306–310. http://doi.org/10.1016/j.bandc.2007.09.008

Heckhausen, H. (1989). *Motivation und Handeln [Motivation and Action]*. Berlin: Springer. http://doi.org/10.1007/978-3-662-08870-8

Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion*, *11*(2), 101–120. http://doi.org/10.1007/BF00992338

Heerwegh, D., & Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in Web Surveys. *Journal of Official Statistics*, *22*(2), 191–210.

Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., Richter, J., & Schreiber, J. (2015). *Studienbereichsspezifische Qualitätssicherung im Bachelorstudium: Befragung der Fakultäts- und Fachbereichsleitungen zum Thema Studienerfolg und Studienabbruch [Study specific quality assurance in Bachelor studies: Survey on academic success and study dro*. Hannover.

Heublein, U., Schmelzer, R., & Sommer, D. (2008). *Die Entwicklung der Studienabbruchquote an den deutschen Hochschulen [Trends of study dropouts at German universities]. HIS: Projektbericht*. Hannover. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Die+Entwicklung+der+Studienabbruchquote+an+den+deutschen+Hochschulen#1

Heydasch, T., Haubrich, J., & Renner, K.-H. (2013). Die Kurzform des Hagener Matrizen-Tests (HMT-S). Ein 6-Item Intelligenztest zum schlussfolgernden Denken [Short version of the Hagener Matrizen-Test (HMT-S): A 6-item intelligence test to measure deductive reasoning].

Methoden, Daten, Analysen, *7*(2), 183–208. http://doi.org/10.12758/mda.2013.011

Hinz, A. M. (1992). Pascal's Triangle and the Tower of Hanoi. *The American Mathematical Monthly*, *99*(6), 538. http://doi.org/10.2307/2324061

Hirst, E., & Schweitzer, M. (1990). Electric-utility resource planning and decision-making: The importance of uncertainty. *Risk Analysis*, *10*(1), 137–146. http://doi.org/10.1111/j.1539-6924.1990.tb01028.x

Hoffman, B. J., Woehr, D. J., Maldagen-Youngjohn, R., & Lyons, B. D. (2011). Great man or great myth? A quantitative review of the relationship between individual differences and leader effectiveness. *Journal of Occupational and Organizational Psychology*, *84*(2), 347–381. http://doi.org/10.1348/096317909X485207

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*(5), 538–540.

Hollen, P. J. (1994). Psychometric properties of two instruments to measure quality decision making. *Research in Nursing & Health*, *17*(2), 137–148. http://doi.org/10.1002/nur.4770170209

Holmes-Rovner, M., Kroll, J., Schmitt, N., Rovner, D. R., Breer, M. L., Rothert, M. L., … Talarczyk, G. (1996). Patient satisfaction with health care decisions: The satisfaction with decision scale. *Medical Decision Making*, *16*(1), 58–64. http://doi.org/10.1177/0272989X9601600114

Horn, J. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements and applications* (pp. 267–300). New York, NY: Wiley.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1991). Feature matching, unique features, and the dynamics of the choice process: Predecision conflict and postdecision satisfaction. *Journal of Experimental Social Psychology*, *27*(5), 411–430. http://doi.org/10.1016/0022-1031(91)90001-M

Howard, R. A. (1966). Decision analysis: Applied decision theory. In *Proceedings of the Fourth International Conference of Operational Research*. Boston, MA.

Howard, R. A. (1968). The foundations of decision analysis. In *Special issue on decision analysis* (pp. 211–219). Op. cit.

Howard, R. A. (1973a). Decision Analsis in Systems Engineering. In R. F. Miles (Ed.), *Systems Concepts: Lectures on Contemporary Approaches to Systems* (pp. 59–93). New York: John Wiley & Sons, Ltd.

Howard, R. A. (1973b). Decision analysis in systems engineering. In R. F. Miles (Ed.), *Systems Concepts: Lectures on Contemporary Approaches to Systems* (pp. 59–93). New York, NY: John Wiley & Sons, Ltd.

Howard, R. A. (1980). An Assessment of Decision Analysis. *Operations Research*, *28*(1), 4–27. http://doi.org/10.1287

Howard, R. A. (2000). Decision in the face of uncertainty. In C. Alexander (Ed.), *Visions of Risk*. London, UK: Pearson Education Limited.

Howard, R. A. (2007). Foundations of decision analysis. In W. Edwards, F. Miles Jr., Raplh, & D. von Winterfeld (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 32–56). New York, NY: Cambridge University Press.

Howard, R. A., & Abbas, A. E. (2015). *Foundations of Decision Analysis* (Global Edi). Essex: Pearson Education Limited.

Hoyle, R. (1995). *Structural equation modeling: Issues and applications*. Newbury Park, CA: SAGE, Publications, Ltd.

Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences*, *10*(1), 31–37. http://doi.org/10.1016/j.tics.2005.11.007

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. http://doi.org/10.1080/10705519909540118

Hudson, D., Seah, L.-H., Hite, D., & Haab, T. (2004). Telephone presurveys, self-selection, and non-response bias to mail and Internet surveys in economic research. *Applied Economics Letters*, *11*(4), 237–240. http://doi.org/10.1080/13504850410001674876

Huitt, W. G. (1992). Problem solving and decision making: Consideration of individual differences using the Myers-Briggs Type Indicator. *Journal of Psychological Type*, *24*, 33–44.

Hunt, R. G., Krzystofiak, F. J., Meindl, J. R., & Yousry, A. M. (1989). Cognitive style and decision making. *Organizational Behavior and Human Decision Processes*, *44*(3), 436–453. http://doi.org/10.1016/0749-5978(89)90018-6

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. http://doi.org/10.1037/0033-2909.96.1.72

IBM Corp. (2012). IBM SPSS Statistics for Mac, (Version 21.0). Armonk, NY: IBM Corp.

Jacobs, J. E., & Klaczynski, P. A. (2005). *The development of judgement and decision making in children and adolescents*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Jacobson, D., Parker, A., Spetzler, C., Bruine de Bruin, W., Hollenbeck, K., Heckerman, D., & Fischhoff, B. (2012). Improved learning in U.S. history and decision competence with decision-focused curriculum. *PLoS ONE*, *7*(9), e45775. http://doi.org/10.1371/journal.pone.0045775

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York, NY: Free Press.

Jones, L. K. (1989). Measuring a three-dimensional construct of career indecision among college students: A revision of the Vocational Decision Scale: The Career Decision Profile. *Journal of Counseling Psychology*, *36*(4), 477–486. http://doi.org/10.1037/0022-0167.36.4.477

Jungermann, H., von Ulardt, I., & Hausmann, L. (1983). The role of the goal for generating action. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analyzing and Aiding Decision Processes*. Amsterdam, NL: Processes, North.

Kahneman, D., & Frederick, S. (2002). Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of Intuitive Ju dgment: Extensions and Applications* (pp. 59–82). New York: Cambridge University Press. Retrieved from https://www.researchgate.net/profile/Shane_Frederick/publication/229071271_Representativeness_revisited_Attribute_substitution_in_intuitive_judgment/links/54087a8c0cf2c48563bd6c75.pdf

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31. http://doi.org/10.1287/mnsc.39.1.17

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342. http://doi.org/10.1111/j.1745-3984.2001.tb01130.x

Keelin, T., Schoemaker, P., & Spetzler, C. (2009). *Decision quality: The fundamentals of making good decisions [Broschure]*. Palo Alto, CA.

Keelin, T., & Spetzler, C. (1992). Decision quality: Opportunity for leadership in total quality management. Palo Alto, CA: Strategic Decisions Group.

Keeney, R. L. (1992). *Value-Focused Thinking: a path to creative decisionmaking*. Cambridge: Harvard University Press.

Keeney, R. L. (1996). Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, *2217*(1992), 4–5. Retrieved from http://www.sciencedirect.com/science/article/pii/0377221796000045

Keeney, R. L. (2004). Making better decision makers. *Decision Analysis*, *1*(4), 193–204. http://doi.org/10.1287/deca.1040.0009

Keeney, R. L. (2007). Developing objectives and attributes. In D. Edwards, Ward; Miles, Raplh, F.; von Winterfeld (Ed.), *Advances in decision analysis: From foundations to applications* (pp. 104–128). New York, NY: Cambridge University Press.

Keeney, R. L. (2008). Personal decisions are the leading cause of death. *Operations Research*, *56*(6), 1335–1347. http://doi.org/10.1287/opre.1080.0588

Keeney, R. L., & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge: Cambridge University Press.

Keller, L. R., & Ho, J. L. (1988). Decision problem structuring: Generating options. *IEEE Transactions on Systems, Man, and Cybernetics*, *18*(5), 715–728. http://doi.org/10.1109/21.21599

Keren, G., & Bruine de Bruin, W. (2003). On the assessment of decision quality: Considerations regarding utility, conflict and accountability. In L. Hardman, David; Macchi (Ed.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*. West Sussex, UK: John Wiley & Sons, Ltd. Retrieved from http://elibrary.kiu.ac.ug:8080/jspui/bitstream/1/1640/1/Thinking Psychological Perspectives on Reasoning,Judgment and Decision Making - David Hardman.pdf#page=349

Kettler, R. J., Braden, J. P., & Beddow, P. A. (2011). Test-taking skills and their impact on accessibility for all students. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of Accessible Achievement Tests for All Students - Bridging the Gaps Between Research, Practice, and Policy* (pp. 147–162). New York, NY: Springer Science & Business Media.

Kim, S., & Hasher, L. (2005). The attraction effect in decision making: Superior performance by older adults. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 120–133. http://doi.org/10.1080/02724980443000160

Kline, P. (2015). *A handbook of test construction: Introduction to psychometric design* (reprint). New York, NY: Routledge.

Kolmogorov, A. N. (2013). *Foundations of the theory of probability*. Eastford, CT, USA: Martino Fine Books.

Krüger, D., Parchmann, I., & Schecker, H. (2013). *Methoden in der naturwissenschaftsdidaktischen Forschung [Methods of educational natural scientific research]*. (H. Schecker, Ed.). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kuhnke, R. (2007). Pretestung des Baseline-Fragebogens und Entwicklung einer Strategie zur Validitätsprüfung von Einzelfragen. [Pre-testing and strategy developement to validate single items]. *DJI Reihe: Wissenschaftliche Texte*. DJI.

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. http://doi.org/10.1037/0022-3514.86.1.148

Kunina, O., Wilhelm, O., Formazin, M., Jonkmann, K., & Schroeders, U. (2007). Extended criteria and predictors in college admission: Exploring the structure of study success and investigating the validity of domain knowledge. *Psychology Science*, *49*(2), 88–114.

Labouvie-Vief, G., DeVoe, M., & Bulka, D. (1989). Speaking about feelings: Conceptions of emotion across the life span. *Psychology and Aging*, *4*(4), 425–437. http://doi.org/10.1037/0882-7974.4.4.425

Lampert, T. (2008). Smoking and passive smoking exposure in young people: Results of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS). *Deutsches Ärzteblatt International*, *105*(15), 265–71. http://doi.org/10.3238/arztebl.2008.0265

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–338). Malden, MA, USA: Blackwell Publishing Ltd.

Larrick, R. P. (2009). Broaden the decision frame to make effective decisions. In E. A. Locke (Ed.), *Handbook of principles of organizational behavior* (2nd ed., pp. 461–480). West Sussex, UK: John Wiley & Sons, Ltd.

Larrick, R. P., Nisbett, R. E., & Morgan, J. N. (1993). Who uses the cost-benefit rules of choice? Implications for the normative status of microeconomic theory. *Organizational Behavior and Human Decision Processes*, *56*(3), 331–347. http://doi.org/10.1006/obhd.1993.1058

Law, K. S., & Mobley, W. M. (1998). Towards a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*(4), 741–755. http://doi.org/10.5465/AMR.1998.1255636

Leventhal, H., Singer, R., & Jones, S. (1965). Effects of fear and specificity of recommendation upon attitudes and behavior. *Journal of Personality and Social Psychology*, *2*(1), 20–29. http://doi.org/10.1037/h0022089

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse [Test construction and analysis]* (6st ed.). Weinheim: Psychologie Verlags Union.

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) [Intelligence-Structure-Test 2000 R]. Göttingen: Hogrefe.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. http://doi.org/10.1006/obhd.1997.2679

Lynch, J. G., Netemeyer, R. G., Spiller, S. A., & Zammit, A. (2010). A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of Consumer Research*, *37*(1), 108–128. http://doi.org/10.1086/649907

MacDonald, G., & Leary, M. R. (2012). Individual differences in self-Esteem. In *Handbook of Self and Identity* (pp. 354–377). New York, NY: Guilford.

Manfreda, K. L., Batagelj, Z., & Vehovar, V. (2006). Design of web survey questionnaires: Three basic experiments. *Journal of Computer-Mediated Communication*, *7*(3), 0–0. http://doi.org/10.1111/j.1083-6101.2002.tb00149.x

Mann, L. (1982). Decision making questionnaires I and II. The Flinders University of South Australia.

Mann, L., Burnett, P., Radford, M., & Ford, S. (1997). The Melbourne Decision Making Questionnaire: an instrument for measuring patterns for coping with decisional conflict. *Journal of Behavioral Decision Making*, *10*(January 1995), 1–19. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-0771(199703)10:1%253C1::AID-BDM242%253E3.0.CO;2-X/abstract

Mann, L., Harmoni, R., & Power, C. (1989). Adolescent decision-making: The development of competence. *Journal of Adolescence*, *12*(3), 265–278. http://doi.org/10.1016/0140-1971(89)90077-8

Mann, L., Harmoni, R., Power, C., Beswick, G., & Ormond, C. (1988). Effectiveness of the GOFER course in decision making for high school students. *Journal of Behavioral Decision Making*, *1*(3), 159–168. http://doi.org/10.1002/bdm.3960010304

Mann, L., Harmoni, R., & Power, C. N. (1988a). GOFER: Basic principles of decision making. Woden: ACT: Curriculum Development Centre.

Mann, L., Harmoni, R., & Power, C. N. (1988b). GOFER: Decision making in practice. Woden: ACT: Curriculum Development Centre.

Mann, L., Radford, M., Burnett, P., Ford, S., Bond, M., Leung, K., … Yang, K.-S. (1998). Cross-cultural Differences in Self-reported Decision-making Style and Confidence. *INTERNATIONAL JOURNAL OF PSYCHOLOGY*, *33*(5), 325–335.

Mason, D. M., & Schuenemeyer, J. H. (1983). A modified Kolmogorov-Smirnov Test sensitive to tail alternatives. *The Annals of Statistics*, *11*(3), 933–946. Retrieved from http://www.jstor.org/stable/2240655

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78. http://doi.org/10.1080/01621459.1951.10500769

Mather, M., & Carstensen, L. L. (2003). Aging and attentional biases for emotional faces. *Psychological Science*, *14*(5), 409–415. http://doi.org/10.1111/1467-9280.01455

Matheson, D., & Matheson, J. E. (1998). *The smart organization: Creating value through strategic R&D*. Menlo Park, CA: Harvard Business School Press.

Matheson, J. E., & Howard, R. A. (1968). *Decision Analysis*. Menlo Park, CA: SRI - Stanford Research Institute.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: a meta-analysis. *Personal Psychology*, *60*(1), 63–91.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

McNamee, P., & Celona, J. (2005). *Decision analysis for the professional* (4th ed.). Menlo Park, CA: SmartOrg, Inc.

Mellers, B., & Locke, C. (2007). What have we learned from our mistakes? In D. Edwards, Ward; Miles, Raplh F.; von Winterfeld (Ed.), *Advances in decision analysis: From foundations to applications* (pp. 351–374). New York, NY: Cambridge University Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. http://doi.org/10.1037/0003-066X.50.9.741

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved?

Perspectives on Psychological Science, *4*(4), 379–383. http://doi.org/10.1111/j.1745-6924.2009.01142.x

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of “unstructured” decision processes. *Administrative Science Quarterly*. Retrieved from http://www.jstor.org/stable/2392045

Montgomery, H. (1988). From cognition to action: The search for dominance in decision-making. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision-making* (pp. 471–483). New York, NY: Wiley.

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. (A. Kelava, Ed.) (2nd ed.). Berlin: Springer.

Muijtjens, A. M. M., van Mameren, H., Hoogenboom, R. J. I., Evers, J. L. H., & van der Vleuten, C. P. M. (1999). The effect of a “don”t know’ option on test scores: Number-right and formula scoring compared. *Medical Education*, *33*(4), 267–275. http://doi.org/10.1046/j.1365-2923.1999.00292.x

MySQL Editions. (n.d.). MySQL [website]. Retrieved September 18, 2015, from https://www.mysql.com/products/

Neisser, U., Boodoo, G., Bouchard, Thomas J., J., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. http://doi.org/10.1037/0003-066X.51.2.77

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling prodecures: Issues and applications*. Thousand Oaks, CA: SAGE, Publications, Inc.

Neubauer, A. C. (2005). Persönlichkeitsunterschiede im Bereich der Fähigkeiten und Kompetenzen – Intelligenz [Individual differences in skills and competences - inteligence]. In H. Weber & T. Rammsayr (Eds.), *Handbuch der Persönlichkeitspsychologie und Differentiellen Psychologie [Handbook of personality and differential psychology]* (pp. 321–332). Göttingen: Hogrefe Verlag GmbH & Co. KG.

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, *58*(2), 367–408. http://doi.org/10.1111/j.1744-6570.2005.00515.x

Nielson, J. (n.d.). Checkboxes vs. radio buttons [Blog post]. Retrieved September 24, 2004, from https://www.nngroup.com/articles/checkboxes-vs-radio-buttons/

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. http://doi.org/10.1037/0033-295X.84.3.231

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. http://doi.org/10.1016/0022-2496(66)90002-2

Nunnally, J. C. (1967). *Psychometric theory* (1st ed.). New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. A. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

Nunnally, J. C., Bernstein, I. A., & ten Berge, J. M. F. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

O’Connor, A., & O’Brien-Pallas, L. L. (1989). Decisional conflict. In G. K. McFarland & E. A. McFarlane (Eds.), *Nursing Diagnosis and Intervention*. Toronto, Ontario, Canada: C.V. Mosby.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3–20). Westport, CT, US: Ablex Publishing.

Orbell, S., & Sheeran, P. (1998). “Inclined abstainers”: A problem for predicting health-related behaviour. *British Journal of Social Psychology*, *37*(2), 151–165. http://doi.org/10.1111/j.2044-8309.1998.tb01162.x

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science*, *23*(5), 381–387. http://doi.org/10.1177/0963721414547414

Oxford Dictionary Online. (n.d.). Oxford Dictionary Online. Retrieved February 12, 2016, from http://www.oxforddictionaries.com/definition/english/

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. http://doi.org/10.1037/0882-7974.17.2.299

Parker, A. M., Bruine de Bruin, W., & Fischhoff, B. (2015). Negative decision outcomes are more common among people with lower decision-making competence: An item-level analysis of the Decision Outcome Inventory (DOI). *Frontiers in Psychology*, *6*(April), 363. http://doi.org/10.3389/fpsyg.2015.00363

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, *18*(1), 1–27. http://doi.org/10.1002/bdm.481

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality Assessment via Questionnaires* (pp. 143–165). Berlin: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-70751-3_8

Payne, S. L. (1951). *The art of asking questions*. Princeton University Press.

Peters, E., Hess, T. M., Västfjäll, D., & Auman, C. (2007). Adult Age Differences in Dual Information Processes: Implications for the Role of Affective and Deliberative Processes in Older Adults' Decision Making. *Perspectives on Psychological Science*, *2*(1), 1–23. http://doi.org/10.1111/j.1745-6916.2007.00025.x

Peters, E., Vastfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and Decision Making. *Psychological Science*, *17*(5), 407–413. http://doi.org/10.1111/j.1467-9280.2006.01720.x

Peterson, M. (2009). *An introduction to decision theory*. Cambridge University Press.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*(2), 381. http://doi.org/10.1086/209405

Phillips, L. D. (2007). Decision conferencing. In W. Edwards, R. F. Miles Jr., & D. von Winterfeldt (Eds.), *Advances in decision analysis: From foundations to applications*. New York, NY: Cambridge University Press.

Phillips, L. D., & Bana e Costa, C. A. (2007). Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*, *154*(1), 51–68. http://doi.org/10.1007/s10479-007-0183-3

Phillips, L. D., & Phillips, M. C. (1993). Facilitated work groups: Theory and practice. *Journal of the Operational Research Society*, *44*(6), 533–549. http://doi.org/10.1038/sj/jors/0440601

Pike, E. O. J. (1973). *Influence of a Test-Taking Skills Instructional Program on the metropolitan achievement tests performance of children from low-income families.* University of Arizona, Tucson, AZ.

Pitz, G. F., Sachs, N. J., & Heerboth, J. (1980). Procedures for eliciting choices in the analysis of individual decisions. *Organizational Behavior and Human Performance*, *26*(3), 396–408. http://doi.org/10.1016/0030-5073(80)90075-6

Porst, R. (2011). *Fragebogen [Surveys]* (3rd ed.). Wiesbaden: VS Verlag.

Preckel, F., & Thiemann, H. (2003). Online- versus paper-pencil-version of a high potential intelligence test. *Swiss Journal of Psychology*, *62*(2), 131–138. http://doi.org/10.1024//1421-0185.62.2.131

Prüfer, P., & Rexroth, M. (1996). Verfahren zur Evaluation von Survey-Fragen. Ein Überblick [Procedures to evaluate survey items - an overview]. *ZUMA-Nachrichten 39*, 95–115.

Raiffa, H. (1968a). *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, MA: Addison-Wesley.

Raiffa, H. (1968b). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Oxford, UK: Addison-Wesley.

Raven, J. C. (2000). Psychometrics, cognitive ability, and occupational performance. *Review of Psychology*, *7*(1–2), 51–74.

Raven, J. C., Raven, J. C., & Court, J. H. (1962). Advanced progressive matrices. London, UK: HK Lewis.

Raven, J. C., Raven, J. C., & Court, J. H. (2003). Manual for Raven's Progressive Matrices and vocabulary scales. San Antonio, TX: Harcourt Assessment.

Reips, U.-D. (2001). The Web Experimental Psychology Lab: Five years of data collection on the

Internet. *Behavior Research Methods, Instruments, & Computers*, *33*(2), 201–211. http://doi.org/10.3758/BF03195366

Reips, U.-D. (2002a). Standards for Internet-Based Experimenting. *Experimental Psychology*, *49*(4), 243–256. http://doi.org/10.1026//1618-3169.49.4.243

Reips, U.-D. (2002b). Theory and techniques of conducting web experiments. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 229–250). Seattle, WA: Hogrefe & Huber.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for fheory, practice, and public policy. *Psychological Science in the Public Interest*, *7*(1), 1–44. http://doi.org/10.1111/j.1529-1006.2006.00026.x

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. http://doi.org/10.1037/a0017327

Rietz, I., & Wahl, S. (1999). Vergleich von Selbst- und Fremdbild von Psychologinnen im Internet und auf Papier. [Comparison of self perception and perception by others of psychologists online and on paper]. In B. Batinic, A. Werner, L. Graef, & W. Bandilla (Eds.), *Online research, methoden, anwendungen, und ergebnisse [Online research, methods, applications, and results]* (pp. 77–92). Göttingen: Hogrefe.

Robson, C. (2011). *Real world research* (3rd ed.). West Sussex, UK: John Wiley & Sons, Ltd.

Rogers, W. T., & Harley, D. (1999). An empirical comparison of three-and four-choice items and tests: Susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement*, *59*(2), 234–247. http://doi.org/10.1177/00131649921969820

Rönnlund, M., Karlsson, E., Laggnäs, E., Larsson, L., & Lindström, T. (2005). Risky decision making across three arenas of choice: Are younger and older adults differently susceptible to framing effects? *The Journal of General Psychology*, *132*(1), 81–93. http://doi.org/10.3200/GENP.132.1.81-93

Rosenberg, M. (1979). *Conceiving the self*. New York, NY: Basic Books.

Ross, J. A. (1981). Improving adolescent decision—making skills. *Curriculum Inquiry*, *11*(3), 279–295. http://doi.org/10.1080/03626784.1981.11075260

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Sabers, D. (1975). Test taking skills. Tucson, AZ: Arizona Center for Educational Research and Development, University of Arizona.

Sáez de Heredia, R. A., Arocena, F. L., & Gárate, J. V. (2004). Decision-making patterns, conflict sytles, and self-esteem. *Psicothema*, *16*(1), 110–116.

Sainfort, F., & Booske, B. C. (2000). Measuring post-decision satisfaction. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *20*(1), 51–61.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*(3), 573–605. http://doi.org/10.1111/j.1744-6570.2003.tb00751.x

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*(4), 140–144. http://doi.org/10.1111/j.0963-7214.2004.00293.x

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equating Models: Test of Significants and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.509.4258&rep=rep1&type=pdf

Schiff, R., & Vakil, E. (2015). Age differences in cognitive skill learning, retention and transfer: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences*, *39*, 164–171. http://doi.org/10.1016/j.lindif.2015.03.010

Schilling, M. S., Oeser, N., & Schaub, C. (2007). How effective are decision analyses? Assessing decision process and group alignment effects. *Decision Analysis*, *4*(4), 227–242. http://doi.org/10.1287/deca.1070.0101

Schlaifer, R. (1959). *Probability and statistics for business decisions: An introduction to managerial economics under uncertainty*. Michigan: McGraw-Hill.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. http://doi.org/10.1037/1040-3590.8.4.350

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll Model of Intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed., pp. 99–144). New York, NY: Guilford Press.

Schnell, R., Hill, P., & Esser, E. (2008). *Methoden der empirischen Sozialforschung [Methods of empirical scocial science]* (8th ed.). München: Oldenbourg.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginners guide to Structure Equating Modeling* (3rd ed.). New York, NY: Taylor and Francis Group, LLC.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Cauals and control beliefs* (pp. 35–37). Windsor, UK: NFER-NELSON.

Scott, S. G., & Bruce, R. a. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818–831. http://doi.org/10.1177/0013164495055005017

Selart, M., & Johansen, S. T. (2011). Understanding the role of Value-Focused Thinking in idea management. *Creativity and Innovation Management*, *20*(3), 196–206. http://doi.org/10.1111/j.1467-8691.2011.00602.x

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591. http://doi.org/10.2307/2333709

Shaughnessy, M. (2009). *Focus in high school mathematics: Statistics and probability*. National Council of Teachers of Mathematics.

Sheeran, P. (2002). Intention—behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, *12*(1), 1–36. http://doi.org/10.1080/14792772143000003

Siebert, J., & Keeney, R. L. (2013). *Stimulating the creation of more and better alternatives* (Wirtschaftswissenschaftliche Diskussionspapiere No. 07–13). Bayreuth.

Siebert, J., & Keeney, R. L. (2015). Creating more and better alternatives for decisions using objectives. *Operations Research*, *63*(5), 1144–1158. http://doi.org/10.1287/opre.2015.1411

Siebert, J., & Kunz, R. (2016). Developing and validating the multidimensional proactive decision-making scale. *European Journal of Operational Research*, *249*(3), 864–877. http://doi.org/10.1016/j.ejor.2015.06.066

Simon, H. A. (1978). Rationality as process and as product of thought. *The American Economic Review*, *68*(2), 1–16. Retrieved from http://www.jstor.org/stable/1816653

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 112–141. http://doi.org/10.1177/0002716212458082

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. http://doi.org/10.1037/0033-2909.119.1.3

Slovic, P. (1962). Convergent validation of risk taking measures. *The Journal of Abnormal and Social Psychology*, *65*(1), 68–71. http://doi.org/10.1037/h0048048

Slovic, P. (2001). Cigarette smokers: Rational actors or rational fools? In P. Slovic (Ed.), *Smoking: Risk, Perception, and Policy* (pp. 97–124). Thousand Oaks, CA: SAGE, Publications, Inc.

Smith, E. R., & Mackie, D. M. (2007). *Social psychology* (3rd ed.). Hove: Psychology Press.

Sniehotta, F. F., Scholz, U., & Schwarzer, R. (2006). Action plans and coping plans for physical exercise: A longitudinal intervention study in cardiac rehabilitation. *British Journal of Health Psychology*, *11*(1), 23–37. http://doi.org/10.1348/135910705X43804

Sniehotta, F. F., Schwarzer, R., Scholz, U., & Schüz, B. (2005). Action planning and coping planning for long-term lifestyle change: theory and assessment. *European Journal of Social Psychology*, *35*(4), 565–576. http://doi.org/10.1002/ejsp.258

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. http://doi.org/10.1037/0278-7393.30.2.299

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of*

Mathematical and Statistical Psychology, *28*(2), 138–151. http://doi.org/10.1111/j.2044-8317.1975.tb00558.x

Spetzler, C. S., & Staël von Holstein, C.-A. S. (1975). Exceptional paper—probability encoding in decision analysis. *Management Science*, *22*(3), 340–358. http://doi.org/10.1287/mnsc.22.3.340

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford, UK: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *The Behavioral and Brain Sciences*, *23*(5), 645–726. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11301544

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695. http://doi.org/10.1037/0022-3514.94.4.672

Stanton, J. M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, *51*(3), 709–725. http://doi.org/10.1111/j.1744-6570.1998.tb00259.x

Statistisches Bundesamt. (2013). *Verkehrsunfälle: Unfälle von 18- bis 24-jährigen im Straßenverkehr (Statistical Report No. 5462406127004) [Traffic accidents: road accidents of adolescents between 18 - 24 years]*. Wiesbaden. Retrieved from www.destatis.de

Stevens, S. S. (1946). *On the theory of scales of measurement*.

Streiner, D. L. (2003). Being Inconsistent About Consistency: When Coefficient Alpha Does and Doesn't Matter. *Journal of Personality Assessment*, *80*(3), 217–222. http://doi.org/10.1207/S15327752JPA8003_01

Suehring, S. (2002). *MySQL Bible*. Hungry Minds Inc.

Szanto, K., de Bruin, W. B., Parker, A. M., Hallquist, M. N., Vanyukov, P. M., & Dombrovski, A. Y. (2015). Decision-making competence and attempted suicide. *The Journal of Clinical Psychiatry*, e1590–e1597. http://doi.org/10.4088/JCP.15m09778

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson Education Inc.

Takemura, K. (2014). *Behavioral decision theory: Psychological and mathematical descriptions of human choice behavior*. Tokyo: Spinger Japan.

Tancredi, L. (1982). Competency for informed consent. *International Journal of Law and Psychiatry*, *5*(1), 51–63. http://doi.org/10.1016/0160-2527(82)90014-0

Tentori, K., Osherson, D., Hasher, L., & May, C. (2001). Wisdom and aging: Irrational preferences in college students but not older adults. *Cognition*, *81*(3), B87–B96. http://doi.org/10.1016/S0010-0277(01)00137-8

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven & London: Yale University Press.

Thunholm, P. (2004). Decision-making style: habit, style or both? *Personality and Individual Differences*, *36*(4), 931–944. http://doi.org/10.1016/S0191- 8869(03)00162-4

Todd, P. M., & Gigerenzer, G. (2000). Précis of Simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(5), S0140525X00003447. http://doi.org/10.1017/S0140525X00003447

Tuten, T. L., Urban, D. J., & Bosnjak, M. (2002). Internet surveys and data quality: A review. In B. Batinic, U.-D. Reips, M. Bosnjak, & A. Werner (Eds.), *Online social sciences* (p. 432). Göttingen: Hogrefe Publishing.

Uhly, A. (2015). *Vorzeitige Vertragslösungen und Ausbildungsverlauf in der dualen Berufsausbildung: Forschungsstand, Datenlage und Analysemöglichkeiten auf Basis der Berufsbildungsstatistik [Premature rescission of contract and training history of the German apprenticeship*. Bonn.

VandenBos, G. R. (2007). *APA dictionary of psychology*. (G. R. VandenBos, Ed.). Washington, DC: American Psychological Association.

Vernon, P. A., & Strudensky, S. (1988). Relationships between problem-solving and intelligence. *Intelligence*, *12*(4), 435–453. http://doi.org/10.1016/0160-2896(88)90006-2

Vogt, W. P. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences.* Newbury Park, CA: SAGE, Publications, Ltd.

von Winterfeld, D., & Edwards, W. (2007). Defining a decision analytic structure. In W. Edwards, R. F. Miles Jr., & D. von Winterfeldt (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 81–103). New York, NY: Cambridge University Press. http://doi.org/10.1017/CBO9780511611308.007

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. http://doi.org/10.1002/bdm.414

Weber, M., & Borcherding, K. (1993). Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research*, *67*(1), 1–12. http://doi.org/10.1016/0377-2217(93)90318-H

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins Co. http://doi.org/10.1037/10020-000

Wechsler, D. (2008). Wechsler adult intelligence scale-fourth. San Antonio, TX: Pearson.

Weller, J. a., Moholy, M., Bossard, E., & Levin, I. P. (2014). Preadolescent decision-making competence predicts interpersonal strengths and difficulties: A 2-year prospective study. *Journal of Behavioral Decision Making*, n/a-n/a. http://doi.org/10.1002/bdm.1822

Whitehead, J. C. (1991). Environmental interest group behavior and self-selection bias in contingent valuation mail surveys. *Growth and Change*, *22*(1), 10–20. http://doi.org/10.1111/j.1468-2257.1991.tb00538.x

Wise, S. L., & Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, *18*(2), 163–183. http://doi.org/10.1207/s15324818ame1802_2

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. University of California, Los Angeles. Retrieved from http://ww.statmodel2.com/download/Yudissertation.pdf

Zedec, S. (2014). APA Dictionary of Statistics and Research Methods. In S. Zedec (Ed.), (p. 448). Washington: American Psychological Association. http://doi.org/10.1037/14336.000

Ziegelmann, J. P., Lippke, S., & Schwarzer, R. (2006). Adoption and maintenance of physical activity: Planning interventions in young, middle-aged, and older adults. *Psychology & Health*, *21*(2), 145–163. http://doi.org/10.1080/1476832050018891

Zimbardo, P. G., & Gerrig, R. J. (1996). *Psychologie [Psychology]* (7th ed.). Berlin: Springer Berlin Heidelberg.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωH: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. http://doi.org/10.1007/s11336-003-0974-7

9 APPENDIX

The following sections and sub-sections present contributory documents and partially statistical details of the present research. For contributory documents the original formatting and layouts are retained. In section 9.1 the pilot study to create the master list for the ability to envision one's objectives is displayed. Section 9.2 shows the evaluation questionnaire of the qualitative pre-testing, following by section 9.3 with the guideline for the semi-structured interviews of the pre-testing. The acquisition emails for the pre-testing are presented for the LSE in section 9.4, for the FU and the UB in section 9.5, and for the main testing for the LSE in section 9.6. The cover letter of the mailing-list for the main testing is displayed in section 9.7. Section 9.8 and its sub-section include the final version of the DAC-test with the complete set of instructions, items and response sets. While section 9.9 presents the DAC-test certificate, which test participants received after the completion of the DAC-test, section 9.10 gives an overview of the surveyed socio-demographics and the test barometer. The chosen measure for validating the DAC-test are listed in section 9.11 and its sub-sections and section 9.12 presents additional results of the item analyses ran in the main testing.

9.1 SURVEYING OBJECTIVES FOR OPERATIONALIZING THE ABILITY TO ENVISION ONE'S OBJECTIVES

In 2012, 59 students of the London School of Economics and Political Science summer school course in "Judgement and Decision Making for Management" helped to generate the master list of objectives for operationalizing the ability to envision one's objectives respectively to the original study of Bond et al. (2008). The main aim for creating such a master list was to provide future test participants an exhaustive list of relevant possible objectives.

The students were on average 20 years old (*M* = 20.21; *SD* = 2.34; range: 18 – 24). Fifty-seven per cent of the group were female and 43 per cent were male. On the very first day of their summer school course, the students were given the following instruction:

Imagine you have to decide what to do after finishing school. What would be your most relevant objectives for choosing a direction? Please list as many objectives as you can think of.

The students had 20 minutes to complete the task. They were informed that this survey is connected to their course work, that they will learn more about how to envision one's objectives in the course, and that the study serves to support a Ph.D. research project.

After eliminating redundancies and aggregating similar objectives to a broader category of objectives, a list of 37 objectives remained (cf. Table 9-1).

Table 9-1 First draft of master list of 37 objectives surveyed in pilot study

Objectives	
1. Acquiring knowledge about something	2. Becoming influential
3. Being independent of parents	4. Being with friends/partner/family
5. Bringing about a positive change in something	6. Characteristics of desired job/profession
7. Contributing positively to society	8. Costs of further education
9. Developing skills	10. Diversity of people to work/spend time with
11. Doing something I am good at	12. Doing something that is important to me
13. Earning my own money	14. Enhancing my resume
15. Enjoying myself	16. Expected work-load
17. Exploring new fields	18. Financial support available
19. Friendliness of environment	20. Geographic location
21. Getting to know the world	22. Having an enjoyable life in the future
23. Having free time	24. Improving chances of finding a job
25. Improving personal status	26. Making new friends/finding a mate
27. Making the world a better place	28. Not having to go to school anymore
29. Personal development	30. Possibility of having fun/party
31. Potential for personal growth	32. Rank of institution (university, company, organization)
33. Receiving support by others	34. Reputation of choice
35. Satisfying parents/family	36. Self-discovery/identification process
37. Serving others, who need help	

The first draft of the master list was presented to a small group of eight psychology master students from Free University of Berlin. Their feedback was used to revise the draft and create the final master list with 43 objectives (cf. appendix sub-section 9.8.1).

9.2 EVALUATION QUESTIONNAIRE OF QUALITATIVE PRE-TESTING

Thank you very much for participating in the pilot-study of the DAC-test.

We would very much appreciate if you could provide us with detailed feedback on your test experience.

In the following you will find a list of expressions that characterise different moods. Please take a look at the list, word by word, and mark for each word the answer that represents best the actual intensity of your mood status. Please judge only how you feel at this moment, and not how you normally or sometimes feel.

Right now I feel …	Definitely not	Not	Not really	A little	Very much	Extreme-ly
... content	○	○	○	○	○	○
... rested	○	○	○	○	○	○
... restless	○	○	○	○	○	○
... bad	○	○	○	○	○	○
... worn-out	○	○	○	○	○	○
... composed	○	○	○	○	○	○
... tired	○	○	○	○	○	○
... great	○	○	○	○	○	○
... uneasy	○	○	○	○	○	○
... energetic	○	○	○	○	○	○
... uncomfortable	○	○	○	○	○	○
... relaxed	○	○	○	○	○	○
... activated	○	○	○	○	○	○
... superb	○	○	○	○	○	○
... calm	○	○	○	○	○	○

Please indicate, for each statement how strongly you agree or disagree.

Taking the test …	Strongly disagree		…		Strongly agree
… was fun	○	○	○	○	○
… was interesting	○	○	○	○	○
… was entertaining	○	○	○	○	○
… was exhausting	○	○	○	○	○
… was boring	○	○	○	○	○
... was diversified	○	○	○	○	○
... gave me a hard time.	○	○	○	○	○
I would like to learn more about individual decision-making.	○	○	○	○	○

I have learned something new taking this test.	○	○	○	○	○

If you did not like some of the tasks, please tell us which and why not?

____

____

____

Please indicate for each statement how strongly you agree or disagree.

	Strongly disagree		...		Strongly agree
The test instruction was displayed well and understandable.	○	○	○	○	○
After reading the instruction I knew what to do and expect.	○	○	○	○	○
The instructions for each item were clear and understandable.	○	○	○	○	○
I always knew what was expected by each item.	○	○	○	○	○
On average the items were too easy.	○	○	○	○	○
On average the items were too difficult.	○	○	○	○	○

If you have perceived (an) item(s) as delicate, please tell us which and why?

____

____

____

If you see any (ethical) issues with the test, please tell us which and why?

____

____

____

If you have ideas that could help to increase the test's professionalism, please tell us here:

____

____

____

What did you like best about the test?

____

____

____

What didn't you like about the test?

____

____

____

Do you still have any remaining/open questions regarding the test?

____

____

____

Thank you very much for your participation and your honesty!

9.3 SEMI-STRUCTURED INTERVIEW GUIDELINE FOR THE DAC PRE-TESTING

→ Annotations for the interviewer are italicized.

A. INTRODUCTION PART

- ***Welcome the interviewee:*** Thank you very much for your time and support!

- ***Explanation of the process of the interview (objective, duration):*** This interview is intended to clarify outstanding issues concerning your test-taking experience and your feedback in the questionnaire. So it will help us to optimise the DAC-test. It will take no longer than 20min.

- ***Question of consent to the recording on tape:*** Do you agree that I can record our conversation on tape? This serves only for simplicity.

- ***Opportunity for questions:*** Do you have any questions before we start?

B. INFORMATION COLLECTION

- First of all, please tell me in a few words how you were feeling during and after having taken the test?

→ Check the answers to item 1 for undesirable answers. If undesirable answers have been given, ask:

- Can you explain why you have perceived the test as...

- Which case (Investment vs. Education) did you like better and why?

- Did you find some of the tasks better or worse than others?

→ If yes, please ask

- Do you remember which and why?

- Do you have any ideas for improvement to address your critique?

→ Check whether the feedback and/or the certificate were assessed negatively. If this appears, ask:

- I can see that you do not like ... What don't you like about it?

- How do you like the test's overall layout?

- Do you have ideas for improvements concerning the layout?

- How do you like the test/item instructions?

- Do you have ideas for improvement concerning the instructions?

- Which items do you like best?

- Which items didn't you like and why?

- Did you perceive items as too easy?

→ If yes, please ask:

- Do you remember which and why?

- Did you perceive items as too difficult?

→ If yes, please ask:

- Do you remember which and why?

→ Check whether items 7 and/or 8 have been filled out (delicate items, ethical issues). If yes, please ask:

- I can see that you perceived items as ... / see any (ethical) issues. Do you remember which and why?

→ Check whether items 9 – 12 have been filled in. If yes, please ask:

- I can see that you ... Can you please explain again what you mean?

- Do you have any general ideas to improve the test?

C. INTERVIEW COMPLETION

- Opportunity for ideas/additions: Would you like to add anything?

- ***Thank the participant:*** Thank you very much for the conversation!

9.4 ACQUISITION E-MAIL FOR PRE-TESTING (LSE)

Dear students,

As part of your course MG110 we invite you to participate in the following test: www.idmc-test.net to assess your individual decision-making competence. The test is **compulsory** and the **deadline** for submission is the **17th of July** 2014.

Duration

Please note that the test will take approx. **2hours** to complete. So please allow yourself enough time. Once you have started with the testing you **cannot** stop and continue later. Of course, you can take a break, but you have to **leave the browser window open**. (If you close it, you will have to start again from the beginning.)

Benefits

By taking the test:

- ...you will receive individual feedback on your performance and discover how capable you are in making decisions.
- ...you will receive a personalised certificate with your test score.

Proving your participation

As the test is anonymous and does not save any personal data, please bring your printed personalised certificate to the course on Friday, 18th of July 2014 to show to me.

About the test

The present IDMC-test is part of the Ph.D. research of Nadine Oeser at LSE. The aim of the research project is to provide a valid and reliable psychometric performance test to measure Individual Decision-Making Competence. The present test is a prototype. Your participation will help to improve its validity.

If you have any clarification questions, please feel free to contact the survey manager: N.Oeser@lse.ac.uk.

Thank you in advance for doing the test.

Best regards,

Gilberto Montibeller

9.5 ACQUISITION E-MAIL FOR PRE-TESTING (FU & UB)

Online study -

Measuring Individual Decision-Making Competence

Thank you for your interest in our research! We are conducting a psychometric test on measuring Individual Decision-Making Competence[84] (IDMC).

Why take the test?

Taking the IDMC-test offers several benefits for you. By completing the IDMC-test...

...you will receive individual feedback on your performance and discover how capable you are in making decisions.

...you will receive a personalised certificate with your test score.

What to expect?

The IDMC consists of different components:

- Agreement to participate, socio-demographic data, pre test barometer (5min)
- Performance test for IDMC (approx. 45min)
- Self-assessment questionnaire (15min)
- Post test barometer and feedback (3min)
- Reasoning matrices test (10min)

All together the test will take you approx. 1 – 1.5 hours to complete. We did our best to make it interesting, entertaining and educating at the same time. Thus, it will be good investment of your time!

Which kinds of aids are allowed?

During the test you are encouraged to use paper, pencil and, calculator.

Attention!

- Our research results depend on high quality data. Thus, please make sure to **follow all test instructions precisely**.
- Please note that the test will take **approx. 1 – 1.5 hours to complete**. So please allow yourself enough time.

[84] Decisions influence the way we live our lives and people face thousands of decisions every day. Many of those decisions are fast and frugal ones, such as what to eat or to drink, or which movie to watch. But obviously, not all decisions people have to address are 'no-brainers'. The majority of important decisions are demanding ones, which are defined by high stakes and serious consequences. e.g. health decisions or educational choices. In those cases the decision environment is complex, uncertain, dynamic, competitive and resources are finite. The competence to solve these kinds of decisions is called Individual Decision-Making Competence (IDMC).

- If you can take a break, you have to **leave the browser window open**. (If you close it, you will have to start again from the beginning.)
- You **cannot go back in the test**. (If you try, the system will force you to start again.)
- On the page you receive your feedback you will be ask to follow a link to a second (very short) test. This is the reasoning matrices test and thus the final part of your testing.

Please select the link to participate in the study —> www.idmc-test.net

Thank you very much in advance!

9.6 ACQUISITION E-MAIL FOR MAIN TESTING (LSE)

Dear students,

Only one week is left until we will welcome you to our LSE summer school this year. We are looking forward to introducing you to the world of decision-making.

In preparation for your course MG110 we would like to test your current decision-making capabilities via an online performance test.

The majority of important decisions that we face are demanding ones, which are defined by high stakes and serious consequences e.g. health decisions or educational choices. In those cases the decision environment is complex, uncertain, dynamic, competitive and resources are finite. The competence to solve these kinds of decisions is called Individual Decision-Making Competence.

The online performance test will allow you to:

- get an initial overview of what complex decision-making is about by being led through the various decision-making tasks of the test
- receive feedback on your individual decision-making performance directly at the end of the test in the form of your test scores
- receive a personalised certificate with your test score at the end of the test

Please follow this link to the test → www.idmc-test.net

About the test

The present IDMC-test is part of the Ph.D. research of Nadine Oeser at LSE, which is supervised by Drs Fasolo & Montibeller. The aim of the research project is to provide a valid and reliable psychometric performance test to measure Individual Decision-Making Competence. With your participation you help to improve its validity.

Deadline

The test is conducted as a preparation for your LSE summer school course. Thus, the deadline for submission is the 5th of July 2015.

Hints

- Please make sure to follow all test instructions precisely.
- The test will take approx. 1.5 hours to complete. So please allow yourself enough time and answer each item with consideration and very thoroughly.
- If you take a break, please leave the browser window open. If you close it, you will have to start over again.

- You cannot go back in the test. If you press enter the system will guide you to the next page.
- On the page with your feedback you will be asked to follow a red link to a second (very short) test. This is the reasoning matrices test and the final part of your testing.

If you have any clarification questions, please feel free to contact the survey manager: N.Oeser@lse.ac.uk.

Many thanks for completing the test. See you soon in London!

Best regards,

Gilberto Montibeller and Barbara Fasolo

9.7 COVER LETTER OF MAILING-LIST FOR MAIN TESTING

Dear all,

My name is Nadine Oeser. I am a Ph.D. student of London School of Economics and Political Science. In the framework of my research I have created a psychometric online performance test to measure Individual Decision-Making Competence for complex decision situations.

After the phase of item analysis and validation I am looking for students (between 18 and 30) for my main testing.

It would be great if you could share my request and the attached handout with your colleagues from decision science or forward the handout to interested students.

Furthermore, it would be great if you could assign participation as homework of your class. In this case I would provide you with a report on the decision competence of your students.

For teachers, lecturers and professors of decision science a performance test measuring Decision-Analytical Competence (DMC) offers three **advantages** (cf. figure):

- First, it would allow pre-post-tests to be conducted. Thus, class teachers, lecturers and co. would be able to test students before and after a course in decision-making or related subjects and compare the results of the two testing times. The results might be used to empirically confirm that you and your course have a positive impact on the Decision-Making Competence of your students.
- Second, it would allow an individual's DMC to be assessed on various dimensions. Hence, a person interested in his/her test score could receive detailed feedback about which dimensions of DMC (e.g. ability to deal with uncertainty or integrate information) are already well evolved and which dimensions need more training.
- Third, it would allow differentiation between people. For example, when recruiting a new employee with good decision-making competence, one could run the test and choose the candidate with the best testing results.

For test participants it would be beneficial to take the test as they...

- ...will receive feedback on their individual decision-making performance
- ...will receive a personalized certificate with their test score

Thank you very much for your support!

Best wishes, Nadine

9.8 THE FINAL DAC-TEST

The following sub-section presents the DAC-test in its final version as it was given to the test participants of the main testing. The original online display of the DAC-test may differ in comparison to this presentation. Page breaks are used so that the actual page-layout is comprehensible.

Sub-section 9.8.1 gives an overview of the task to envision one's objectives. Sub-section 9.8.2 shows the list of the seven items for the ability to realise decision opportunities and sub-section 9.8.3 the eight items to measure the ability to assess decision fitness. The eight items of the ability to frame a decision are shown by sub-section 9.8.4 and the seven items of the ability to identify relevant alternatives by sub-section 9.8.5. The remaining seven items of the ability to deal with uncertainty are presented in sub-section 9.8.6. Sub-section 9.8.7 gives an overview of the 12 items of the ability to integrate information.

For each item an explicit item ID, for instance EO_01, is assigned so that the statistical results presented within this research can clearly be assigned to the corresponding item. However, the item ID are not presented to the test participants.

9.8.1 Ability to Envision One's Objectives

Imagine you have to decide what to do after finishing school. What would be your most relevant objectives for choosing a direction? Please list as many objectives as you can think of, writing each one in the lines (from A to AD) below (cf. example item).

A
B
C
D
E
...
....
...
Z
AA
AB
AC
AD

Again, imagine after you finished school you have to decide what to do now with your life. Please select all objectives that appear relevant to you for selecting a direction for your life by ticking the checkbox on the left of each objective in the list below (cf. example item).

I would like to ...

☐ ... acquire knowledge about a specific topic.
☐ ... be independent of parents/earn my own money.
☐ ... contribute positively to society in the short- or long-run.
☐ ... develop/improve my skills.
☐ ... do something for which I could get financial support.
☐ ... do something I am good at.
☐ ... do something in a desirable geographic location.
☐ ... do something that is challenging.
☐ ... do something that is diversified.
☐ ... do something that is fun.
☐ ... do something that allows me to keep my pet/take care of my pet.
☐ ... do something that allows me to take care of my grandma, younger brother, etc.
☐ ... do something that fulfils criteria like helping humans or animals, dealing with IT, saving our planet, being creative, etc.
☐ ... do something that gives me planning reliability/security.
☐ ... do something that gives me the chance to be with my family/friends/partner.
☐ ... do something that gives me the chance to make new friends/find a mate, etc.
☐ ... do something that gives me the chance to travel/live abroad/get to know the world, etc.
☐ ... do something that gives me the choice of self-discovery/identification process.
☐ ... do something that improves my personal status in the short- or long-run.
☐ ... do something that is accessible without any problems with my disability.
☐ ... do something that is affordable/not to expensive/well paid.
☐ ... do something that is family-friendly.
☐ ... do something that is important to me.
☐ ... do something that paves the way for becoming influential.
☐ ... do something that paves the way for a high salary.
☐ ... do something that paves the way for good career prospects.
☐ ... do something that paves the way for having an enjoyable life in the future.
☐ ... do something that provides a good work-life-balance.
☐ ... do something that takes place in a friendly environment where people support each other.
☐ ... do something with a good reputation/which enhances my resume.
☐ ... do something which is more practically or more theoretically oriented.
☐ ... do something which satisfies my parents/others.
☐ ... enjoy myself/celebrate/party, etc.
☐ ... explore new fields/do research, etc.
☐ ... get in touch with diverse people.
☐ ... improve my chances of finding a job.
☐ ... meet people who are like I am.
☐ ... personally grow from what I do.
☐ ... take on responsibility.
☐ ... try something totally new.

Please match each objective you listed in the first task (now displayed on the left side below) to the objectives on the right side by writing its letter to the left of them. If some of your personal objectives do NOT match any objective here, please write them down in the shaded area below (cf. example item).

I would like to ...

___ ☐ ... acquire knowledge about a specific topic.
___ ☐ ... be independent of parents/earn my own money.
___ ☐ ... contribute positively to society in the short- or long-run.
___ ☐ ... develop/improve my skills.
___ ☐ ... do something for which I could get financial support.
___ ☐ ... do something I am good at.
___ ☐ ... do something in a desirable geographic location.
___ ☐ ... do something that is challenging.
___ ☐ ... do something that is diversified.
___ ☐ ... do something that is fun.
___ ☐ ... do something that allows me to keep my pet/take care of my pet.
___ ☐ ... do something that allows me to take care of my grandma, younger brother, etc.
___ ☐ ... do something that fulfils criteria like helping humans or animals, dealing with IT, saving our planet, being creative, etc.
___ ☐ ... do something that gives me planning reliability/security.
___ ☐ ... do something that gives me the chance to be with my family/friends/partner.
___ ☐ ... do something that gives me the chance to make new friends/find a mate, etc.
___ ☐ ... do something that gives me the chance to travel/live abroad/get to know the world, etc.
___ ☐ ... do something that gives me the choice of self-discovery/identification process.
___ ☐ ... do something that improves my personal status in the short- or long-run.
___ ☐ ... do something that is accessible without any problems with my disability.
___ ☐ ... do something that is affordable/not to expensive/well paid.
___ ☐ ... do something that is family-friendly.
___ ☐ ... do something that is important to me.
___ ☐ ... do something that paves the way for becoming influential.
___ ☐ ... do something that paves the way for a high salary.
___ ☐ ... do something that paves the way for good career prospects.
___ ☐ ... do something that paves the way for having an enjoyable life in the future.
___ ☐ ... do something that provides a good work-life-balance.
___ ☐ ... do something that takes place in a friendly environment where people support each other.
___ ☐ ... do something with a good reputation/which enhances my resume.
___ ☐ ... do something which is more practically or more theoretically oriented.
___ ☐ ... do something which satisfies my parents/others.
___ ☐ ... enjoy myself/celebrate/party, etc.
___ ☐ ... explore new fields/do research, etc.
___ ☐ ... get in touch with diverse people.
___ ☐ ... improve my chances of finding a job.
___ ☐ ... meet people who are like I am.
___ ☐ ... personally grow from what I do.
___ ☐ ... take on responsibility.
___ ☐ ... try something totally new.

Finally look at all objectives you have selected in the second task and the ones in the shaded area. Please mark the 10 most relevant ones for you by ticking the checkbox on their very left side. Don't worry if you have selected less than 10 objectives (cf. example item).

I would like to ...

☐ ___ ☐ ... acquire knowledge about a specific topic.
☐ ___ ☐ ... be independent of parents/earn my own money.
☐ ___ ☐ ... contribute positively to society in the short- or long-run.
☐ ___ ☐ ... develop/improve my skills.
☐ ___ ☐ ... do something for which I could get financial support.
☐ ___ ☐ ... do something I am good at.
☐ ___ ☐ ... do something in a desirable geographic location.
☐ ___ ☐ ... do something that is challenging.
☐ ___ ☐ ... do something that is diversified.
☐ ___ ☐ ... do something that is fun.
☐ ___ ☐ ... do something that allows me to keep my pet/take care of my pet.
☐ ___ ☐ ... do something that allows me to take care of my grandma, younger brother, etc.
☐ ___ ☐ ... do something that fulfils criteria like helping humans or animals, dealing with IT, saving our planet, being creative, etc.
☐ ___ ☐ ... do something that gives me planning reliability/security.
☐ ___ ☐ ... do something that gives me the chance to be with my family/friends/partner.
☐ ___ ☐ ... do something that gives me the chance to make new friends/find a mate, etc.
☐ ___ ☐ ... do something that gives me the chance to travel/live abroad/get to know the world, etc.
☐ ___ ☐ ... do something that gives me the choice of self-discovery/identification process.
☐ ___ ☐ ... do something that improves my personal status in the short- or long-run.
☐ ___ ☐ ... do something that is accessible without any problems with my disability.
☐ ___ ☐ ... do something that is affordable/not to expensive/well paid.
☐ ___ ☐ ... do something that is family-friendly.
☐ ___ ☐ ... do something that is important to me.
☐ ___ ☐ ... do something that paves the way for becoming influential.
☐ ___ ☐ ... do something that paves the way for a high salary.
☐ ___ ☐ ... do something that paves the way for good career prospects.
☐ ___ ☐ ... do something that paves the way for having an enjoyable life in the future.
☐ ___ ☐ ... do something that provides a good work-life-balance.
☐ ___ ☐ ... do something that takes place in a friendly environment where people support each other.
☐ ___ ☐ ... do something with a good reputation/which enhances my resume.
☐ ___ ☐ ... do something which is more practically or more theoretically oriented.
☐ ___ ☐ ... do something which satisfies my parents/others.
☐ ___ ☐ ... enjoy myself/celebrate/party, etc.
☐ ___ ☐ ... explore new fields/do research, etc.
☐ ___ ☐ ... get in touch with diverse people.
☐ ___ ☐ ... improve my chances of finding a job.
☐ ___ ☐ ... meet people who are like I am.
☐ ___ ☐ ... personally grow from what I do.
☐ ___ ☐ ... take on responsibility.
☐ ___ ☐ ... try something totally new.

☐ ____________________
☐ ____________________
☐ ____________________
☐ ____________________
☐ ____________________
☐ ____________________
☐ ____________________

☐ I have selected 10 objectives (max).

9.8.2 Ability to Realise Decision Opportunities

Please read the following case descriptions carefully and decide whether the character is facing a concrete decision situation or not by ticking the appropriate button.

If you think that the character is facing a decision situation, an additional question will appear. If this is the case please assess the potential impact of this decision on the character's life by ticking the appropriate button (cf. scale explanation).

If everything goes well, Luca will obtain his school-leaving qualifications next year. He has talked a lot with his parents about what he wants to become when he is grown up. Now he is thinking about what to do after passing his exams. Tomorrow, he will be able to apply for study programmes online, using his last school report.

DO_01	○ No	○ Yes		○ I do not know
DO_02	○ Big decision	○	○ Small decision	○ I do not know

Inès is delighted. She has received an offer to enrol in highly sought-after further education, which is offered by the company she is working for. The seminar is starting next May and will last for 6 months. Actually, she wanted to go on holiday in August. She has already booked the flights to Madeira. However, Inès knows that if she enrols she won't be allowed to take vacations.

DO_05	○ No	○ Yes		○ I do not know
DO_06	○ Big decision	○	○ Small decision	○ I do not know

Kilian is studying at the Montgomery University. Right now Kilian is taking a multiple-choice exam. The question he is facing just in this second is quite difficult. He is fairly sure, that two of the answers given are unlikely to be true. But for the other two, he does not know at all.

DO_07	○ No	○ Yes		○ I do not know
DO_08	○ Big decision	○	○ Small decision	○ I do not know

Last week Antoine and his friends had a meeting with their career advisor in order to develop some ideas about what to do after finishing school. Antoine has tried to find out more about possible courses of study. His advisor tells him that for pursuing a medical degree, James would need much better grades.

DO_09	○ No	○ Yes		○ I do not know
	○ Big decision	○	○ Small decision	○ I do not know

Two months ago, Christopher started dating a girl from his university. He had secretly fancied her since he had first seen her on campus. During a friend's party he had the courage to talk to her and they had fun dancing together. She became Christopher's first girlfriend. It was a very exciting time for him and he spent almost all his free time with her. Today, Christopher is facing his first year final exams. This morning on his way to university he feels he should have spent more time learning.

DO_10	○ No	○ Yes		○ I do not know
	○ Big decision	○	○ Small decision	○ I do not know

Conan is in his third semester of studying Chemistry at university. He enjoys being in university a lot. He has many friends, a new kind of freedom and exciting experiences in the huge city of London. Up to now he has been doing quite well, but recently, his marks have got worse. There is just too much going on to be able to concentrate on the learning part of studying. Today, Conan is told by his professor that he will not pass the course on Physical Chemistry this semester. Conan exclaims: "But Sir, if you won't let me pass this would mean I cannot go on studying!"

DO_12	○ No	○ Yes		○ I do not know
	○ Big decision	○	○ Small decision	○ I do not know

9.8.3 Ability to Assess Decision Fitness

Please read the following case descriptions carefully and pay attention to the character's physical condition and emotions. Please decide whether it is advisable to make a decision in his or her situation. For each case, please tick the appropriate button (cf. scale explanation).

Susan's parents expect her to become a dentist, like her father. Susan however would rather study literature. Tonight they had a heated argument on this topic. Her parents refuse to finance her "foolish ideas" and Susan shouted back at them. She ran back to her room, slammed the door and thought: "I will show them how I can study literature. I don't need their help! I will complete my online application now."

DF_01	O No	O Yes	O I do not know

Thomas is in the middle of his preparations for his final examination. He is studying the whole day, but it is very difficult for him to concentrate. He always thinks about what to do after his exam. "Should he study? Should he take a Gap Year and travel? Or maybe it would be better to do an internship first?" It is only one week until he has to be fit for his exam and he really needs to concentrate on his studies. "If I decide now on what to do after my exam, I don't need to think about it any longer and I will be able to truly concentrate on my exam", he thinks.

DF_02	O No	O Yes	O I do not know

Sophie is working at the bar of a club in order to make some money for traveling. After leaving school, she did not know what to do and so she decided to take some sort of Gap Year first. She wanted to earn some money, do a bit of travelling and maybe an internship or two. This morning she came home at 7.30 in the morning because there was a big party at the club she is working in. Completely exhausted, she fell into her bed. At 9 am, she is suddenly awoken by her telephone ringing. A friend of hers has a surprising idea. She proposes to fly to India together. Sophie's friend has found a very cheap offer for flight tickets that she could book for the two of them. Unfortunately, the offer will expire today so she wants Sophie to take the decision immediately.

DF_03	O No	O Yes	O I do not know

Daniel has just finished his final exams in high school. He decided to take up his studies in molecular biology. He has come to this decision after an internship in a research institute during the last summer break. He asked the researchers which universities they recommend for studying and also visited some of the universities in his country to collect his own impressions. He compared the education costs and found out that there were ways to get financial support. Now he has offers from three different programmes of study and the deadline for registration is tomorrow.

DF_04	○ No	○ Yes	○ I do not know

Cheryl is enraged. She just talked to her boss about the inappropriate behaviour of her colleague. "He is making remarks about my dress and my figure and he is making whistling noises when I pass. He doesn't stop even though I have told him to stop it." Her boss didn't seem to take her seriously. "He is just making compliments to his good-looking colleague. Other women would be happy if they elicit this kind of response from a charming man." At first, she was so stunned, she didn't know what to say. She looked at her boss and thought "He cannot be serious!" She left his office, slammed the door and said aloud. "This is it, I am quitting this job!"

DF_05	○ No	○ Yes	○ I do not know

Nasir is 24 and has just had a job interview. The day was exhausting; he had to get up at 5 am to be at the site on time. Then he had to demonstrate his analytical thinking in a computer task that took 2.5 hours. After a little snack, he went through an assessment centre with interviews, a group discussion, role plays and an "in tray" task. At the end of the day, he had a feedback round. All in all, he must have mastered the tasks to the interviewers' satisfaction. They offered him a job and said he could sign the contract right away if he wished so. Nasir cannot believe his luck. The good news almost cleared his exhaustion away. He feels a warm tingling in his stomach and a sense of elation.

DF_06	○ No	○ Yes	○ I do not know

Mrs Baker has been a surgeon for 15 years now. She likes her job very much. As a cardiothoracic surgeon, she is performing surgery on people who have serious problems with their heart or lungs. The operations are usually very long (2 - 5 hours) and demanding. Mrs Baker has to be highly concentrated and work to within a millimetre. Lately however, her job has become more and more challenging. Her back is hurting and her fingers are tingling after standing in OR for a couple of hours. She is less concentrated and has trouble sleeping. Last week during a cardiopulmonary bypass operation, she even had to ask her assistant surgeon to take over. Now she has requested a week of unpaid leave in order to recover and think about her options. She will not get any younger and the work will remain demanding. Maybe she should hand in her notice and become a general practitioner?

DF_07	○ No	○ Yes	○ I do not know

Christian has always wanted to become an engineer in the automotive industry. However, when he applied for admission to study mechanical or automotive engineering, he was rejected by each university. His grades didn't fully meet the requirements and there seemed to be many applications. Christian decided to re-apply the following year. To get to learn more about the work and processes he started working as an unskilled production helper in a car factory. The year passed and again his applications were rejected. Christian kept working his shifts in the factory, annoyed by the fact that he is not able to work there as an engineer and that there is still no place at university in sight. He wonders whether his application will be more successful one year later. As he was always interested in architecture, too, he could also see himself studying this. Christian researches on the Internet and procures all brochures he needs and finds out that his chances of being accepted for architecture right away are much higher.

DF_08	○ No	○ Yes	○ I do not know

9.8.4 Ability to Frame a Decision

Please, read the following decision case carefully.

Martine currently attends a college in the UK. She will finish her final exams in school next summer. She has always been a comparatively good student. Now she and her classmates have to start thinking about what to do after obtaining their school leaving certificates. Last Wednesday they had an information session at school where a former student talked about his experiences during a Gap Year. When Martine came home that day, she told her parents about what she had heard about the Gap Year but her parents were not very enthusiastic. “Maybe you should better think about what you are going to do with your life. You can’t spend your life just travelling and expecting us to pay for you. Unfortunately, we would not be able to support you financially.”
The conversation with her parents made Martine seriously think about what she’d do after leaving school. She feels inclined towards educating or supporting young people. Maybe she should become a teacher or social worker!? However, for becoming either one, she would have to take up studies at some university. Applications would take place soon, with her next grade report. Hopefully her grades would be good enough to be accepted.
Later that day, Martine called her friend Patricia to complain about her parents’ reaction:
Martine: “Hi Trish! Imagine, my parents won’t let me do a Gap Year. They say I should earn my own money first. But I really would love to travel the world or at least live in another county for some time.”
Patricia: “Oh no, and what are you going to do now?”
Martine: “I don’t know. Maybe I could do a Gap Year with work and travel and study afterwards. That would give me the possibility to see the world without being dependent on financial support. But I think that won’t give me the possibility to save money for my study afterwards.”
Patricia: “Yeah, I guess you’re right.”
Martine: “Are you going to study after finishing school?”
Patricia: “Yes, I’ll study Business and Economics. The companies always need managers.”
Martine: “Hmm. I’d like to become a teacher or a social worker. Do you think it will be expensive?”
Patricia: “Well, if your parents cannot support you it will be hard. Fees are quite high. But you could take up a student loan.”
Martine: “No, I definitely do not want to get indebted! Maybe I can work part-time.”
Patricia: “Yes, but that might not be enough to finance your study. You should try to get a scholarship! That would solve your financial problems, allow you to study right away and those scholarship programmes often include a semester abroad.”
Martine: “Sounds like a possibility. I will talk to Professor Morgan about that tomorrow. Thanks and ‘night Trish”
Patricia: “Good night. See you tomorrow.”
Just before midnight Martine went to bed but could not sleep for quite a while. “Is there a likelihood to get a scholarship? How high is the chance of finding enough jobs during the Gap Year, which allow me to earn more money than I’ll be spending? Maybe I should follow my parents’ advice and earn my own money first. After working for one or two years I would have saved some money to study. That is a quite appealing idea. But in this case, I probably will not be able to see a lot of the world.” ...

While Martine and Patricia are having lunch in the school canteen the next day, Patricia is making some notes on what Martine is telling her.

Please have a look at the following list of facts Patricia has collected, and decide which kind of information it is in the decision-making situation by choosing the most appropriate description from the drop-down menu for each fact <u>(cf. definitions)</u>.

Item ID	Fact	Drop-down menu with response set
FD_01	Saving money and travelling now	Objectives Conflicting objectives Alternative(s) Uncertainty Time frame Decision-maker Resources Friend Family Location
FD_02	Grades	
FD_03	Would love to travel the world or at least live in another county for some time and educating and/or supporting young people	
	Martine	
FD_04	Supporting young people and saving money	
FD_05	Go straight to university and find a job to earn money	
	Application period is close	
FD_06	Take a gap year and study then	
FD_07	Try to get a scholarship	
	Patricia	
FD_08	Getting a scholarship	
	Martine’s parents	

9.8.5 Ability to Identify Relevant Alternatives

Please read the following decision problem carefully:

Brad, a classmate of Martine is about to complete his secondary education as well. He is also facing the question of what to do after school. “Well, what alternatives do I have for now?” he asks himself.

Here is what Brad thinks is relevant regarding his decision:

- “I am really happy to be out of school now. I am not going to study straight away – Forget it!”
- “Whatever it is – I need to get some money for living. My parents aren’t going to pay.”
- “I have always liked the countryside. Which alternatives are out there that let me spend some time outside?”

Knowing more about Brad’s objectives now, please choose which of the following alternatives fit his three objectives all at the same time. Please tick the appropriate buttons (cf. scale explanation).

IA_03	Work on a fishing boat	○ Does not fit	○ Does fit	○ I do not know
IA_04	Volunteering in an orphanage	○ Does not fit	○ Does fit	○ I do not know
IA_05	Starting an apprenticeship as a carpenter	○ Does not fit	○ Does fit	○ I do not know
IA_06	Going on an outdoor holiday	○ Does not fit	○ Does fit	○ I do not know
IA_07	Go on a pilgrimage	○ Does not fit	○ Does fit	○ I do not know
IA_08	Start online-gaming	○ Does not fit	○ Does fit	○ I do not know
IA_09	Become a photo-journalist	○ Does not fit	○ Does fit	○ I do not know
IA_10	Win a scholarship and study geography	○ Does not fit	○ Does fit	○ I do not know

9.8.6 Ability to Deal with Uncertainty

Martine thought a lot about her alternatives. Then she has the idea to write each alternative on a piece of paper and just draw one of them. In this case, destiny would decide for her. “But is it just luck to win a lottery?” She asks herself.

Please read the following case description carefully.

While thinking about this question, Martine decides to actually take part in a lottery. There are two different lotteries: one of them says: “Every 20th lottery ticket is a winning ticket!” Hmmm… this makes 5%, thinks Martine. She knows of another lottery with a probability of winning which is 10%.

Which lottery should Martine play to maximise her expected value? Please tick the appropriate button

DU_01	○ 5%	○ 10%	○ I do not know

Please read the following case description carefully.

Martine really has to focus, so she thinks about two scholarships her teacher has told her about. They would be a good chance for her. The association offering the two scholarships (A and B) takes affirmative action in order to promote equality of opportunity for members of both sexes (50/50). Therefore, the applicants are selected to keep the proportion of men and women equal. To be able to assess which of the scholarships

	Scholarship A (20 participants)	**Scholarship B (30 participants)**
Men	**34**	**85**
Women	**80**	**113**

offers the higher probability for being selected, she requests a list of applicants (cf. figure):

Applicants are chosen randomly by a computer programme. What is the probability of Martine being selected for scholarship A and for scholarship B if the proportion of men and women should be equal? Please give your answer by selecting the closest whole number from the lists below.

DU_02

The probability of Martine being selected for scholarship A is

○ 0%	○ 1%	○ 2%	○ 3%	○ 5%
○ 10%	○ 13%	○ 17%	○ 20%	○ 35%
○ 52%	○ 80%	○ 90%	○ 95%	○ 96%

○ 97%	○ 98%	○ 99%	○ 100%

DU_03

The probability of Martine being selected for scholarship B is

○ 0%	○ 1%	○ 2%	○ 3%	○ 5%
○ 10%	○ 13%	○ 17%	○ 20%	○ 35%
○ 52%	○ 80%	○ 90%	○ 95%	○ 96%
○ 97%	○ 98%	○ 99%	○ 100%	

Martine decides to apply for scholarship A. What would a probability of 17% mean in her case? Please tick the appropriate button.

DU_04

○ Martine will definitely get the scholarship.

○ If she applied for the same scholarship 100 times, she would be admitted on exactly 17 occasions.

○ Martine might get a scholarship.

○ Out of 17 persons who applied, one will get a scholarship.

○ Martine will not get a scholarship.

Please read the following case description carefully.

As the next big school holidays are not far away, Martine's friend Patricia has applied for an internship with a big company in London to gain some work experience. This company has published some statistics on its website:

- Of all applicants, 15% are invited to interviews.

- Of those who had the chance to present themselves, 20% are offered a position.

What is the probability of getting a job offer after having sent application materials? Please select the right answer by choosing the closest whole number from the list below.

DU_05

○ 0%	○ 1%	○ 2%	○ 3%	○ 5%
○ 10%	○ 13%	○ 17%	○ 20%	○ 35%
○ 52%	○ 80%	○ 90%	○ 95%	○ 96%
○ 97%	○ 98%	○ 99%	○ 100%	

Let's say the probability of getting this job is 3%. Patricia thinks: "Wow, that is not really a good chance of getting this job. I could also try to roll doubles with 6-sided dice. What would be more likely?" Please calculate the probability of getting doubles with a 6-sided dice and select the right answer by choosing the closest whole number from the list below.

DU_06

○ 0%	○ 1%	○ 2%	○ 3%	○ 5%
○ 10%	○ 13%	○ 17%	○ 20%	○ 35%
○ 52%	○ 80%	○ 90%	○ 95%	○ 96%
○ 97%	○ 98%	○ 99%	○ 100%	

Please read the following case description carefully.

Martine has found out more about financial schemes for studying at two different universities.

The first university has annual tuition fees of 9,000£. The only funding available is a National Scholarship – about 200,000 students apply for it every year, but only one tenth will be successful. Winning it will give you an annual grant of 3,000£.

The other university has fees of £8,000 per year. There is a lottery amongst all applicants to distribute 15 full scholarships (these lucky people will not have any fees to pay) and 30 awards of 3,000£. About 500 students apply for this lottery.

What is the expected value (cf. definition) of costs for a three-year program at each university? Please DO NOT ROUND UP OR DOWN (cf. scale explanation).

DU_07	The expected value for the first university is _ _. _ _ £.
DU_08	The expected value for the second university is _ _. _ _ £.

9.8.7 Ability to Integrate Information

Please read the following case description carefully.

Martine is still undecided about what to do after her A-Levels. She is thinking about going abroad to gain some experience. Eventually, she would like to become a teacher. However, her parents are not able to support her financially. She therefore has to find a way of supporting herself.

Let's take a closer look at her objectives, her alternatives and how she should decide.

Her objectives are:

- Does want to have money to study

- Educating and/or supporting young people

- Would love to travel the world or at least live in another country for some time

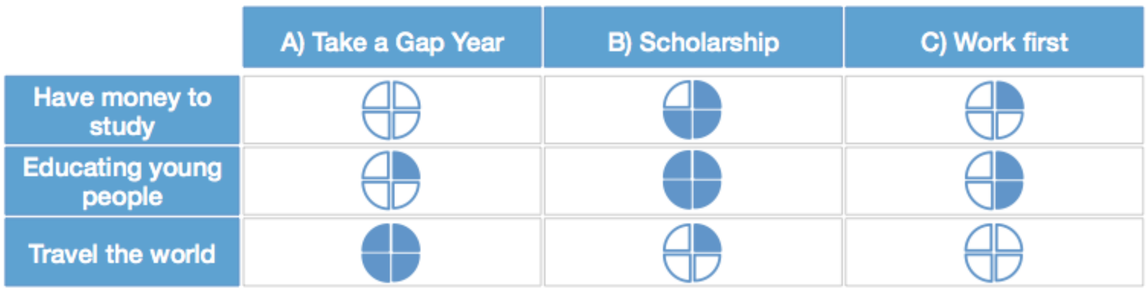
And the alternatives are:

A) Take a Gap Year and study afterwards

B) Go straight to university and get a scholarship

C) Find a job to earn money, save the money and study afterwards

Martine assesses how well her objectives are met by each alternative (cf. figure) by using pie icons:



Please put the alternatives in order according to Martine's objectives for her decision. Start with the most preferred alternative (cf. scale explanation).

II_04	Most preferred alternative		
	○ A	○ B	○ C
	○ A and B	○ A and C	○ B and C
	○ A, B, and C	○ /	
II_05	Second most preferred alternative		
	...		
II_06	Least preferred alternative		
	....		

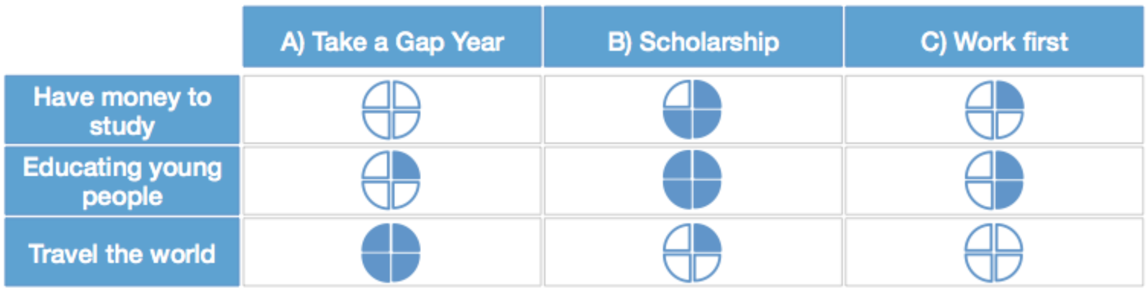
But Martine's objectives aren't all equally important:

- Having money for her study is in between: 33%

- Educating and/or supporting young people is the most relevant: 47%

- Travelling the world/living in another country is the least relevant: 20%

To remember cf. figure:



Please put the alternatives in order according to Martine's assessment of the relevance of her objectives. Start with the most preferred alternative (cf. scale explanation).

II_07 Most preferred alternative

○ A ○ B ○ C

○ A and B ○ A and C ○ B and C

○ A, B, and C ○ /

II_08 Second most preferred alternative

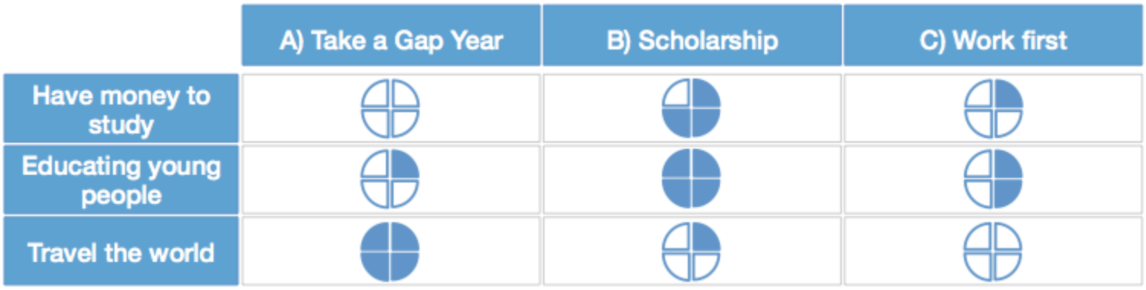
...

II_09 Least preferred alternative

....

It looks like Martine can now easily decide which alternative to choose to best pursue her objectives. But there are certain risks she has not yet included in her calculations.

Her marks might not be good enough for getting a scholarship. In this case studying on a scholarship would not be an option any more. Martine thinks that she has a 20% chance to get the scholarship. (For your calculation please assume that in case of not getting the scholarship the expected value for alternative B would be 0.) To remember cf. figure:



Please put the alternatives in order according to the probabilities of Martine's alternatives (cf. scale explanation).

II_10	Most preferred alternative		
	○ A	○ B	○ C
	○ A and B	○ A and C	○ B and C
	○ A, B, and C	○ /	
II_11	Second most preferred alternative		
	...		
II_12	Least preferred alternative		
	....		

Martine has simplified the comparison of alternatives by distributing pie icons by gut feeling in order to assess how well one alternative fits one objective. It was clever of her to use one type of scale for all objectives, as it makes it easier to compare.

But Martine thinks again about the comparison: Maybe she could have done a more precise job. What if she had used whole numbers between 0 (zero fit) and 100 (perfect fit) to assess how well one alternative fits one objective?

Here is the information on the money Martine could save:

A) Take a Gap Year and study afterwards: £5,000

B) Go straight to university and get a scholarship: £48,000

C) Find a job to earn money, save the money and study afterwards: £15,000

Please translate the money Martine would be able to save for her study, into whole numbers between 0 and 100 for each alternative, given that

- £55,000 is the whole amount of money Martine needs to finance her study

- Martine does not want to get indebted

- the more money the better

- the scale is interval-scaled, which means that the distance between 0 and 20 is equivalent to the distance between 60 and 80.

II_13	Take a Gap Year and study afterwards (£5,000): __ __ (please round to whole numbers)
II_14	Go straight to university and get a scholarship (£48,000): __ __ (please round to whole numbers)
II_15	Find a job to earn money, save the money and study afterwards (£15,000): __ ___ (please round to whole numbers)

9.9 DAC-TEST CERTIFICATE

CERTIFICATE

This is to certify that

First name Surname

has successfully completed

the DAC

a psychometric test to assess
Decision-Analytical Competence

Ability to recognise decision opportunities	80%
Ability to assess decision fitness	80%
Ability to frame a decision	80%
Ability to envision one's objectives	80%
Ability to identify relevant alternatives	80%
Ability to deal with uncertainty	80%
Ability to integrate information	80%
Total	**80%**

London, date (dd/mm/yyyy)

Nadine Oeser

9.10 SOCIO-DEMOGRAPHIC VARIABLES & TEST BAROMETER

Please help us to map your data from the different test parts and enter an alphanumeric code following the instruction below.

1. = The first letter of the street you live in (e.g. Morgan Street → M)
2. = The fifth number of your local phone number (e.g. 33567-32 → 7)
3. = The third letter of your mothers first name (e.g. Maggie → G)
4. = Your favourite number (e.g. My favourite number is 23 → 23)
5. = The first letter of your fathers first name (e.g. Frank → F)

Before the DAC-test

The following few questions are not part of the assessment of your Individual Decision-Making Competence. They just serve to give us some further information.

Before you can start with the testing, please give us some information about your background. This is needed for statistical purposes.

What is your current age?

☐ I confirm that I am 18 years or older.

How would you rate your English language skills?

O Basic English skills O Good/conversant O Fluent

O Very good/business-fluent O Mother tongue

Have you learned about good decision-making before?

O No, I have never heard about it. O Yes, I have learned about good decision-making

If YES:

☐ I have taken a class in school/university about decision analysis or decision science.

☐ I had a workshop on decision-making

☐ I have read a book on the topic

☐ Other → Please describe:

How did you get to know about this test? Please select your source:

☐ LSE ☐ FU ☐ HU ☐ UB ☐ VT ☐ MPI

☐ Misc. → Please specify:

How far do the following statements apply to you? Please indicate, for each statement how strongly you agree or disagree.

I am taking this test because…	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
…I am interested in individual decision-making.	○	○	○	○	○
…I would like to learn more about individual decision-making.	○	○	○	○	○
…I would like to find out about my own decision-making competence.	○	○	○	○	○
…I want to reach a high test score.	○	○	○	○	○
...I would like to find out about my own decision-making competence.	○	○	○	○	○
…I would like to have a certificate of participation.	○	○	○	○	○

Again, how strongly do you agree with each statement given below?

	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
Doing well on this test is important to me.	○	○	○	○	○
I want to be among the top scorers in this test.	○	○	○	○	○
I am determined to complete this test.	○	○	○	○	○
I usually do pretty well on tests.	○	○	○	○	○
My test scores don't usually reflect my true abilities.	○	○	○	○	○
Once I undertake a task, I usually push myself to my limits.	○	○	○	○	○
I try to do well in everything I undertake	○	○	○	○	○

After the DAC-test

Finally, we would like to ask you a few more questions about your background. These are not used for identifying you, but needed for statistical purposes.

Are you...?

○ Male ○ Female ○ Other

What is your nationality?

What is your highest completed degree of education?

O No school leaving certificate
O Certificate after 9 or 10 years of school education
O University-entrance diploma
O Certificate for apprenticeship
O Bachelor's degree
O Diploma or Master's degree
O PhD
O Other → Please state:

What is your mark average of your school leaving examination?

O A+ O A O A- O B+ O B O B-
O C+ O C O C- O D+ O D

What is/was your mark average in mathematics at your school leaving examination?

O A+ O A O A- O B+ O B O B-
O C+ O C O C- O D+ O D

Have you already entered professional life?

O Yes O No

What is your former, current or intended profession? Which field does it belong to?

O Architecture and Engineering Occupations
O Arts, Design, Entertainment, Sports, and Media Occupations
O Building and Grounds Cleaning and Maintenance Occupations
O Business and Financial Operations Occupations
O Community and Social Services Occupations
O Computer and Mathematical Occupations
O Construction and Extraction Occupations
O Education, Training, and Library Occupations
O Farming, Fishing, and Forestry Occupations
O Food Preparation and Serving Related Occupations
O Healthcare Practitioners and Technical Occupations
O Healthcare Support Occupations
O Installation, Maintenance, and Repair Occupations
O Legal Occupations
O Life, Physical, and Social Science Occupations

O Management Occupations

O Military Specific Occupations

O Office and Administrative Support Occupations

O Personal Care and Service Occupations

O Production Occupations

O Protective Service Occupations

O Sales and Related Occupations

O Transportation and Material Moving Occupations"

O Other → Please state:

Do/did you currently/previously have managerial responsibility?

O Yes O No

If YES, how many persons did you supervise?

O 1-5 O 6-20 O 21-50 O 51-100 O More than 100

9.11 CRITERIA FOR VALIDATION – CHOSEN MEASURES

The present appendix-section lists the chosen measures that were applied to validate the DAC-test. To assess decision self-esteem the Melbourne Decision Making Questionnaire I is used (cf. sub-section 9.11.1), to assess differences in decision-making style the Melbourne Decision Making Questionnaire II (cf. sub-section 9.11.2), and to assess decision satisfaction the Satisfaction with Decision Scale (cf. sub-section 9.11.3). One aspect of intolerance of uncertainty is measured by a sub-scale of Uncertainty Impeding Action (cf. sub-section 9.11.4), fluid intelligence is captured by the short version of the Hagen Matrices Test (cf. sub-section 9.11.5), and problem-solving competence is assessed by the Tower of Hanoi (cf. sub-section 9.11.6).

9.11.1 Melbourne Decision Making Questionnaire I – Decision Self-Esteem

Within this Ph.D. research, decision self-esteem is measured by the six items of the Melbourne Decision Making Questionnaire I of Mann et al. (1998). In the following the applied instruction is displayed and Table 9-2 shows its items with the response set:

In the following, you'll find a number of statements on how one might describe oneself. Please tick the appropriate button to describe how accurately each statement describes you. Please answer according to your first impression.

Table 9-2 Melbourne Decision Making Questionnaire I – Decision Self-Esteem (Mann et al., 1998)

#	Item	Scale		
		Not true for me	Sometimes true	True for me
1	I feel confident about my ability to makes decisions.	○	○	○
2	I feel inferior to most people in making decisions. (R)	○	○	○
3	I think that I am a good decision maker.	○	○	○
4	I feel so discouraged that I give up trying to make decisions (R)	○	○	○
5	The decisions I make turn out well.	○	○	○
6	It is easy for other people to convince me that their decision rather then mine is the correct one. (R)	○	○	○

Note. (R) = The bracketed R identifies negatively-keyed items that are phrased so that an agreement with the item represents a relatively low level of the attribute being measured. If a survey contains positively- and negatively-keyed items, the latter item type has to be reverse-scored for statistical analyses.

9.11.2 Melbourne Decision Making Questionnaire II – Decision-Making Style

Various styles of decision-making are assessed by the 22 items of the Melbourne Decision Making Questionnaire II of Mann et al. (1997). Below the scale instruction is presented and Table 9-3 displays the items with the response set:

In the following, you'll find a number of statements on how one might describe oneself. Please tick the appropriate button to describe how accurately each statement describes you. Please answer according to your first impression.

Table 9-3 Melbourne Decision Making Questionnaire II – Decision-Making Style (Mann et al., 1997)

#	sc	Item	Scale		
			Not true for me	Sometimes true	True for me
1	V	I like to consider all the alternatives.	○	○	○
2	V	I try to find out the disadvantages of all alternatives.	○	○	○
3	V	I consider how best to carry out a decision.	○	○	○
4	V	When making decisions I like to collect a lot of information.	○	○	○
5	V	I try to be clear about my objectives before choosing.	○	○	○
6	V	I take a lot of care before choosing.	○	○	○
7	B	I avoid making decisions.	○	○	○
8	B	I do not make decisions unless I really have to.	○	○	○
9	B	I prefer to leave decisions to others.	○	○	○
10	B	I do not like to take responsibility for making decisions.	○	○	○
11	B	If a decision can be made by me or another person I let the other person make it.	○	○	○
12	B	I prefer that people who are better informed decide for me.	○	○	○
13	P	I waste a lot of time on trivial matters before getting to the final decision.	○	○	○
14	P	Even after I have made a decision I delay acting upon it.	○	○	○
15	P	When I have to make a decision I wait a long time before starting to think about it.	○	○	○
16	P	I delay making decisions until it is too late.	○	○	○
17	P	I put off making decisions.	○	○	○
18	H	Whenever I face a difficult decision I feel pessimistic about finding a good solution.	○	○	○
19	H	I feel as if I am under tremendous time pressure when making decisions.	○	○	○
20	H	The possibility that some small thing might go wrong causes me to swing abruptly in my preference.	○	○	○
21	H	I cannot think straight if I have to make a decision in a hurry.	○	○	○
22	H	After a decision is made I spend a lot of time convincing myself it was correct.	○	○	○

Note. sc = sub-scale of MDMQ-II; V = Vigilance; B = Buck-passing; P = Procrastination; H = Hypervigilance.

9.11.3 Satisfaction with Decision Scale

The Satisfaction with Decision Scale of Holmes-Rovner et al. (1996) is used to measure decision satisfaction. The scale instruction is shown and the six items with their corresponding response set are presented in Table 9-4:

Please think about the latest complex decision you were facing. Answer the following questions about your decision. Please indicate to what extent each statement is true for you at this time.

Table 9-4 Satisfaction with Decision Scale (Holmes-Rovner et al., 1996)

#	Item	Scale				
		strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
1	I am satisfied that I am adequately informed about the issues important to my decision.	○	○	○	○	○
2	The decision I made was the best decision possible for me personally.	○	○	○	○	○
3	I am satisfied that my decision was consistent with my personal values.	○	○	○	○	○
4	I expect to successfully carry out (or continue to carry out) the decision I made.	○	○	○	○	○
5	I am satisfied that this was my decision to make.	○	○	○	○	○
6	I am satisfied with my decision.	○	○	○	○	○

9.11.4 Intolerance of Uncertainty – Sub-scale Uncertainty Impeding Action

Buhr and Dugas (2002) developed the Intolerance of Uncertainty Scale, of which the sub-scale *uncertainty leads to the inability to act* is applied for this research. The applied instruction and its eight items with their response set (cf. Table 9-5) are shown in the following:

Please think about the latest complex decision you were facing. Answer the following questions about your decision. Please indicate to what extent each statement is true for you at this time.

Table 9-5 Intolerance of Uncertainty – Sub-scale Uncertainty Impeding Action (Buhr & Dugas, 2002)

#	Item	Scale				
		strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
1	Uncertainty stops me from having a strong opinion.	○	○	○	○	○
2	Uncertainty keeps me from living a full life.	○	○	○	○	○
3	When I am uncertain, I can’t go forward.	○	○	○	○	○
4	Being uncertain means that I am not first rate.	○	○	○	○	○
5	When I am uncertain, I can’t function very well.	○	○	○	○	○
6	The smallest doubt can stop me from acting.	○	○	○	○	○
7	Being uncertain means that I lack confidence.	○	○	○	○	○
8	I must get away from all uncertain situations.	○	○	○	○	○

9.11.5 Hagen Matrices Test – Short Version

To assess fluid intelligence the short version of the Hagen Matrices Test of Heydasch et al. (2013) is applied. Figure 9-1 gives an overview of its instruction and an example item.

Dear participant,

this test is about finding rules in abstract patterns and to complete them in a logical way. Each task shows an incomplete jigsaw puzzle. The patterns you will see follow rules which may apply to a row, a column or to a diagonal. They may apply to the figure as a whole or to parts of it only. They may involve addition, subtraction, the alignment of figures or single components. Only one of the eight pieces given is the correct one required to complete the design.

It is your task to select the piece which completes the jigsaw puzzle. Each task needs to be completed within 2:00 minutes.

First sample
Which piece is the one required to complete the design?
(Click on the correct piece and then on "Continue")

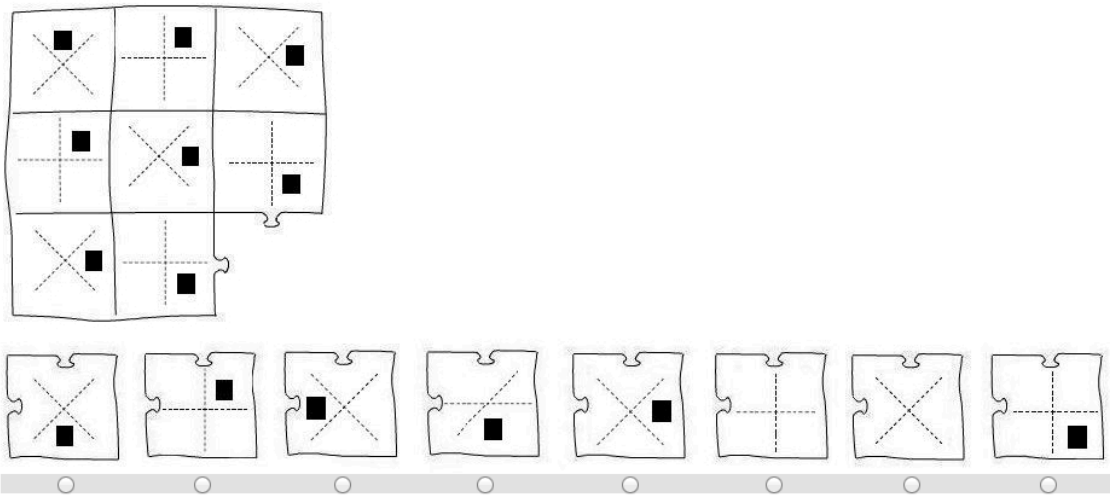


Figure 9-1 Hagen Matrices Test – Short Version (Heydasch et al., 2013)

9.11.6 Problem Solving Competence – Tower of Hanoi

Problem-solving competence is captured by the Tower of Hanoi. As the Tower of Hanoi is a relatively well-known task, whose solution can be found on the Internet, within the present research, the Tower of Hanoi was named ring task. In the following the instruction and a visual clarification (cf. Figure 9-2), which was also presented to the test participants, are shown:

For the following task, please try to place all rings from the middle on one of the other rods. It does not matter whether you choose the left or the right rod.

There are three rules you have to follow:

- A larger ring can never be placed on a smaller ring.

- You can only move one ring at one time.

- A ring can only be moved if it is the topmost ring on a pile.

Try to complete the task in as little moves as possible. One move is started whenever you choose a ring to be moved. A move cannot be made undone. You can see your number of attempted moves below (cf. example).

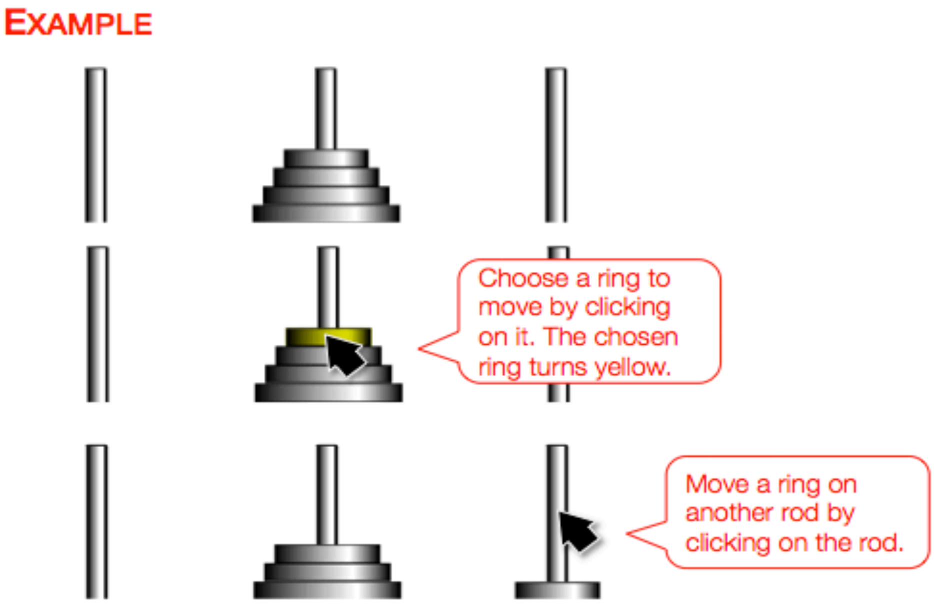


Figure 9-2 Problem Solving Competence – Tower of Hanoi

9.12 ITEM ANALYSIS OF MAIN TESTING

For the sake of completeness, the current appendix section presents the statistical properties for the items of the DAC-test sub-scales of the main testing, including item difficulty indices and discriminatory power coefficients.

Table 9-6 gives an overview of the statistical properties for the items of the DAC-test sub-scale OPPORTUNITIES. Item DO_01 was excluded from the final statistical analyses as it turned out to be unintentionally negatively-poled.

Table 9-6 Statistical properties of OPPORTUNITIES (main testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	AoMT	Reason for exclusion from statistical analyses
DO_01	.77	.42	.78	.56	.37	no	unintentionally negatively-poled
DO_02	.60	.49	.77	.65	.53	yes	
DO_05	.90	.31	.89	.79	.23	yes	
DO_06	.23	.42	.25	-.12	.38	yes	
DO_07	.76	.43	.78	.56	.53	yes	
DO_08	.66	.47	.87	.80	.63	yes	
DO_09	.57	.50	.60	.20	.46	yes	
DO_10	.78	.41	.80	.60	.49	yes	
DO_12	.69	.46	.73	.46	.51	yes	

Note. *N* = 368; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; AoMT = selected for the analyses of the main testing; DO = Item ID for the DAC-dimension of OPPORTUNITIES.

Table 9-7 shows the statistical properties for the items of FITNESS. Item DF_02 was excluded from the final statistical analyses as it turned out to be negatively-poled.

Table 9-7 Statistical properties of FITNESS (main testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	AoMT	Reason for exclusion from statistical analyses
DF_01	.72	.45	.73	.47	.51	yes	
DF_02	.58	.49	.61	.22	.31	no	unintentionally negatively-poled
DF_03	.49	.50	.51	.02	.64	yes	
DF_04	.95	.22	.96	.93	.11	yes	
DF_05	.69	.46	.72	.44	.57	yes	
DF_06	.44	.50	.47	-.06	.57	yes	
DF_07	.77	.42	.81	.62	.30	yes	
DF_08	.80	.40	.87	.74	.34	yes	

Note. *N* = 368; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; AoMT = selected for the analyses of the main testing; DF = Item ID for the DAC-dimension of FITNESS.

The statistical properties for the items of FRAME are presented in Table 9-8.

Table 9-8 Statistical properties of FRAME (main testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	AoMT	Reason for exclusion from statistical analyses
FD_01	.44	.50	.44	.39	.50	yes	
FD_02	.27	.44	.27	.20	.31	yes	
FD_03	.62	.49	.62	.58	.55	yes	
FD_04	.48	.50	.48	.43	.38	yes	
FD_05	.56	.50	.56	.52	.56	yes	
FD_06	.55	.50	.55	.50	.55	yes	
FD_07	.41	.49	.41	.36	.51	yes	
FD_08	.37	.48	.37	.31	.47	yes	

Note. *N* = 368; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; AoMT = selected for the analyses of the main testing; FD = Item ID for the DAC-dimension of FRAME.

The statistical properties for the items of the DAC-test sub-scale ALTERNATIVES are displayed in Table 9-9. Item IA_05 was excluded from the final statistical analyses as it turned out to be unintentionally negatively-poled.

Table 9-9 Statistical properties of ALTERNATIVES (main testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	AoMT	Reason for exclusion from statistical analyses
IA_03	.91	.29	.91	.82	.15	yes	
IA_04	.77	.42	.83	.67	.41	yes	
IA_05	.52	.50	.58	.16	.31	no	unintentionally negatively-poled
IA_06	.76	.43	.79	.59	.54	yes	
IA_07	.66	.47	.72	.44	.67	yes	
IA_08	.89	.31	.94	.88	.20	yes	
IA_09	.81	.39	.85	.69	.39	yes	
IA_10	.78	.41	.80	.60	.39	yes	

Note. *N* = 368; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; AoMT = selected for the analyses of the main testing; IA = Item ID for the DAC-dimension of ALTERNATIVES.

Table 9-10 gives an overview of the statistical properties for the items of the DAC-test sub-scale UNCERTAINTY. Item DU_01 was excluded from the final statistical analyses as it turned out to be unintentionally negatively-poled.

Table 9-10 Statistical properties of UNCERTAINTY (main testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	AoMT	Reason for exclusion from statistical analyses
DU_01	.78	.42	.90	.81	.17	no	unintentionally negatively-poled
DU_02	.61	.49	.61	.59	.81	yes	
DU_03	.57	.50	.57	.54	.82	yes	
DU_04	.55	.50	.55	.44	.47	yes	
DU_05	.81	.39	.82	.81	.40	yes	
DU_06	.43	.50	.43	.40	.50	yes	
DU_07	.20	.40	.22	*	.45	yes	
DU_08	.11	.32	.13	*	.26	yes	

Note. *N* = 368; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; AoMT = selected for the analyses of the main testing; DU = Item ID for the DAC-dimension of UNCERTAINTY.

* Items are not multiple-choice items. Guessing correction is not necessary to calculate the item difficulty index.

Table 9-11 shows the statistical properties for the items of INFORMATION.

Table 9-11 Statistical properties of INFORMATION (main testing)

Item ID	*M*	*SD*	p_i	p_{i_c}	r_{it}	AoMT	Reason for exclusion from statistical analyses
II_04	.82	.38	.82	.80	.42	yes	
II_05	.69	.46	.69	.65	.64	yes	
II_06	.74	.44	.74	.70	.57	yes	
II_07	.86	.35	.86	.84	.37	yes	
II_08	.32	.47	.32	.22	.74	yes	
II_09	.29	.45	.29	.19	.66	yes	
II_10	.28	.45	.28	.18	.60	yes	
II_11	.39	.49	.39	.30	.55	yes	
II_12	.30	.46	.31	.21	.63	yes	
II_13	.50	.50	.54	*	.85	yes	
II_14	.51	.50	.55	*	.84	yes	
II_15	.53	.50	.57	*	.87	yes	

Note. *N* = 368; p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; r_{it} = item discriminatory power coefficient; AoMT = selected for the analyses of the main testing; II = Item ID for the DAC-dimension of INFORMATION.

* Items are not multiple-choice items. Guessing correction is not necessary to calculate the item difficulty index.

Table 9-12 presents the distribution of the item difficulty indices of the pre-testing and the main testing.

Table 9-12 Distribution of item difficulty indices (p_i /p_{i_c}) per DAC-dimension (pre-testing & main testing)

Scale	Pre-testing			Main testing		
	M	SD	*Range*	*M*	SD	*Range*
DO	.53	.20	.23 – .80	.54	.22	.20 – .80
DF	.44	.17	.21 – .76	.53	.29	.02 – .93
FD	.42	.15	.21 - .61	.41	.12	.20 - .58
IA	.45	.15	.22 - .65	.67	.15	.44 - .88
DU	.36	.23	.05 - .72	.45	.23	.13 - .81
II	.40	.21	.20 - .78	.48	.25	.18 - .84

Note. N = 143 (pre-testing); N = 368 (main testing); p_i = item difficulty index/100; p_{i_c} = corrected item difficulty index/100; DO = Ability to recognise decision opportunities; DF = Ability to assess decision fitness; FD = Ability to frame a decision; IA = Ability to identify relevant alternatives; DU = Ability to deal with uncertainty; II = Ability to integrate information.