

DIPLOMARBEIT

VERHALTENSBEOBSACHTUNG IM KONTEXT VON

DYNAMISCHEN GRUPPENPROZESSEN

Operationalisierung und Prüfung der Gütekriterien

Sarah Baeske

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl lt. Studienblatt: A 298

Studienrichtung lt. Studienblatt: Psychologie

Betreuerin / Betreuer: Prof. Dr. Georg Gittler

INHALTSVERZEICHNIS

ABSTRACT	VII
-----------------	-----

ABBILDUNGSVERZEICHNIS	VIII
------------------------------	------

TABELLENVERZEICHNIS	IX
----------------------------	----

GLIEDERUNG

1. GRUNDLEGUNG	1
-----------------------	---

1.1 Einordnung	1
----------------	---

1.2 Der Dyadic Cooperation Test – DCT	2
---------------------------------------	---

1.3 Anforderungsanalysen	5
--------------------------	---

1.4 Unfallstatistik	5
---------------------	---

1.5 Zielsetzung der Diplomarbeit	6
----------------------------------	---

1.6 Gliederung der Diplomarbeit	6
---------------------------------	---

2. THEORIE	7
-------------------	---

2.1 Dimensionen	7
-----------------	---

2.1.1 Belastungsfähigkeit	8
---------------------------	---

2.1.1.1 Physiologische Reaktion	9
---------------------------------	---

2.1.1.2 Subjektive Befindlichkeit	10
-----------------------------------	----

2.1.1.3 Leistungsfähigkeit	10
----------------------------	----

2.1.2 Arbeitsstil	11
-------------------	----

2.1.2.1 Motivationspsychologischer Ansatz	13
---	----

2.1.2.2 Kognitive Stile	15
-------------------------	----

2.1.2.3	Flexibilität	17
2.1.3	Interaktion & Kommunikation	17
2.1.3.1	Reziprozität & Symmetrie vs. Kompensation & Komplementarität	19
2.1.3.2	Kontrolle & Dominanz	20
2.1.3.3	Kooperation, Wettbewerb & Konflikt	20
2.1.3.4	Interpersoneller Einfluss & Konformität	21
2.2	Verhaltensbeobachtung	21
2.2.1	Güte einer ratingbasierten Verhaltensbeobachtung	24
2.2.2	Videovermittelte Verhaltensbeobachtung	25
2.2.3	Verhaltensbeobachtung im Assessment Center	27
2.3	Beobachtertraining	27
2.3.1	Beobachterfehler	28
2.3.1.1	Halo Effekt	30
2.3.1.2	Implizite Persönlichkeitstheorie	30
2.3.1.3	Primacy-/Recency-Fehler	30
2.3.1.4	Bezugsfehler	31
2.3.1.5	Similiar-to-me-Effekt	31
2.3.1.6	Grundeinstellung des Beobachters	31
2.3.1.7	Dominanz einzelner Beobachter	32
2.3.2	Inhalte und Ziele des Beobachtertrainings	32
3.	FRAGESTELLUNG	35

4. METHODIK	36
4.1 Operationalisierung	36
4.1.1 Generierung der Indikatoren	36
4.1.2 Klassifikation der Indikatoren	38
4.1.2.1 Klassifikation Belastbarkeit	39
4.1.2.2 Klassifikation Arbeitsstil	40
4.1.2.3 Klassifikation Interaktion	42
4.1.2.4 Klassifikation kognitive Fähigkeiten	42
4.1.3 Quantifizierung der Indikatoren	43
4.1.3.1 Zuweisungsregeln Belastbarkeit	44
4.1.3.2 Zuweisungsregeln Arbeitshaltung	46
4.1.3.3 Zuweisungsregeln Interaktion	49
4.1.3.4 Zuweisungsregeln kognitive Leistung	51
4.1.4 Bewertungsbogen DCT	52
4.2 Beobachtertraining	53
4.3 Methodik der Datenerhebung	55
4.3.1 Stichprobe	55
4.3.2 Untersuchungsdesign	56
4.3.3 Datenaufbereitung	56
4.3.3.1 Deskriptiv	57
4.3.3.2 Beobachterübereinstimmung	57

4.3.3.3	Beobachterfehler	60
5.	ERGEBNISSE	61
5.1	Deskriptiv	61
5.2	Beobachterübereinstimmung	65
5.3	Beobachterfehler	69
5.4	Beantwortung der Forschungsfragen	71
6.	DISKUSSION UND AUSBLICK	73
6.1	Diskussion der Hypothesen	73
6.1.1	Forschungsfrage 1	73
6.1.2	Forschungsfrage 2	75
6.2	Integrative Diskussion	76
6.3	Ausblick	77
7.	ZUSAMMENFASSUNG	80
	ANHANG	81
	QUELLENVERZEICHNIS	89
	DANKSAGUNG	95
	EIDESSTAATLICHE VERSICHERUNG	96
	LEBENS LAUF	97

ABSTRACT

Requirements analyses and investigations of aviation accidents have pointed out the necessity of non-cognitive skills in different fields of air traffic control. In the context of theoretical observations, like cognitive abilities attributes such as resilience, work-style and social interaction are commonly referred to as crucial indicators of performance. The Dyadic Cooperation Test (DCT) provided the opportunity to assess these dimensions within the selection process of the flight control services of the department for selection of the specialized group flight psychology at the center for aviation and space medicine of the German Air Force. In order to implement this procedure, the theoretical construction had to be operational and subsequently checked for quality criteria. This approach marks the purpose of the thesis at hand. With a great demand for transparency the following steps of construction will be described in detail: video-based generation of indicators, classification of scales and quantification, as well as examination of performance data. It has been shown that the newly developed scales physiological resilience, subjective resilience, impulsiveness, flexibility, communication, dominance and cognitive performance provide objective and therefore psychometrically applicable data. As to the cooperation and frustration tolerance scales, precise adjustments will be necessary for a scientifically substantiated application. A coaching test for supervisors based on newly developed observation training has proven to improve objectivity in most cases. Although additional studies – e.g. to examine validity – will have to follow, this diploma thesis succeeds to show that this revision and extension of the DCT has the potential to be incorporated in the selection process and to optimize applicant selection for the gain of further information.

ABBILDUNGSVERZEICHNIS

Abbildung 1 :	Aufbau des DC	4
Abbildung 2:	Unterscheidung Belastungs-, Beanspruchungs- & Bewertungs- /Folgeanalyse nach Wieland-Eckelmann	13
Abbildung 3:	Leistungsfähigkeit und Belastung nach Litzcke & Schuh	11
Abbildung 4:	Vier Seiten einer Nachricht nach Schulz von Thun	18
Abbildung 5:	Datenschema als kaskadierte Prädikation nach Faßnacht	23
Abbildung 6:	Zusammensetzung Messwert nach Annahme der klassischen Testtheorie	28
Abbildung 7:	Illustration einiger Beobachterfehler nach Amelang & Schmidt - Atzert	32
Abbildung 8:	Codiersystem der qualitativen Inhaltsanalyse des Videomaterials	39
Abbildung 9:	Beobachtungsbogen DCT	53
Abbildung 10:	Beispielfolien Beobachtertraining DCT	55
Abbildung 11:	Untersuchungsdesign	56
Abbildung 12:	Häufigkeitsverteilung der bewerteten Skalen gesamt	62
Abbildung 13:	Häufigkeitsverteilung der bewerteten Skalen vor dem Training	63
Abbildung 14:	Häufigkeitsverteilung der bewerteten Skalen nach dem Training	64
Abbildung 15:	Single Measure vor und nach Beobachterschulung	68
Abbildung 16:	Average Measure vor und nach Beobachterschulung	69

TABELLENVERZEICHNIS

Tabelle 1:	Facetten des Fünf-Faktoren Modells nach McCrae & Costa	13
Tabelle 2:	Kognitive Stile nach Häcker & Stampf	16
Tabelle 3:	Quantifizierung physiologische Belastungsfähigkeit	45
Tabelle 4:	Quantifizierung subjektive Belastungsfähigkeit	46
Tabelle 5:	Quantifizierung Impulsivität vs. Reflexivität	47
Tabelle 6:	Quantifizierung Frustrationstoleranz	48
Tabelle 7:	Quantifizierung Flexibilität	49
Tabelle 8:	Quantifizierung Teamfähigkeit	49
Tabelle 9:	Quantifizierung Dominanz vs. Submissivität	50
Tabelle 10:	Quantifizierung Kommunikation	51
Tabelle 11:	Quantifizierung kognitive Leistung	52
Tabelle 12:	Standardmatrix zur Berechnung der ICC nach von Eye und Mun	59
Tabelle 13:	Datenmatrix zur Berechnung der ICC	60
Tabelle 14:	Personenspezifische Bewertungsanzahl vor und nach Schulung	65
Tabelle 15:	Die Intraclass-Correlation (ICC) der einzelnen Skalen	67
Tabelle 16:	Differenz Mittel Einzelbewertung - Mittel Gesamtbewertung	70

1. GRUNDLEGNUNG

Dieses Kapitel beschäftigt sich mit der Zielsetzung sowie einer thematischen Einordnung der vorliegenden Diplomarbeit. Das grundlegende Verfahren, der Dyadic Cooperation Test, wird bezüglich Aufbau und Ablauf im Detail vorgestellt. Aktuelle Unfallstatistiken und Anforderungsanalysen spannen den theoretischen Rahmen und zeigen die Notwendigkeit der bearbeiteten Fragestellung. Im Anschluss daran folgt ein kurzer Ausblick in Bezug auf Gliederung und Inhalte der folgenden Abschnitte.

1.1 Einordnung

Wie Anforderungsanalysen zeigen und aus Unfallstatistiken abzuleiten ist, sind nicht-kognitive Fähigkeiten, wie Kommunikationsverhalten und Kooperation wichtige Fertigkeiten im Anforderungsprofil von Tätigkeitsfeldern der Luftfahrtindustrie, so auch in der militärischen Flugverkehrskontrolle. Diese Tatsache ist für die Bewerberauswahl der Bundeswehr von großer Bedeutung. Mit der Zielsetzung einer optimierten prognostischen Validität der Eignungsdiagnostik müssen die bisherigen Konstrukte also um weitere Merkmale erweitert werden. Zur Erfassung von Kommunikationsverhalten und Kooperation gibt es derzeit einige psychologische Testverfahren. Ein Beispiel hierfür sind die entsprechenden Items zu „Teambindung“ und „Umgang mit anderen“ der Manager Potential Analyse (MPA) aus dem Wiener Testsystem der Firma Schuhfried. Persönlichkeitsfragebögen wie diese gehen jedoch meist mit der Problematik sozial erwünschter Antworttendenzen einher und sind gerade in Auswahl-situationen eher ungeeignet (Kubinger, 2009, S. 121). Ein valideres Verfahren ist die Verhaltensbeobachtung (Hunter & Martinussen, 2008, S.44). Hierbei wird anhand entsprechender Szenarien das Bewerberverhalten beobachtet und bewertet. Im Anschluss wird der Bewerber hinsichtlich seiner berufsspezifischen Eignung beurteilt. Für möglichst objektive Resultate sollten die Beobachter gezielt geschult und angeleitet werden. Dabei sind allgemeines Vorgehen, der Verweis auf mögliche Beobachterfehler und die zu beobachtenden Dimensionen von zentraler Bedeutung.

Zur Auswahl der Flugführungsdienste¹ erhebt das Dezernat Auswahl der Fachgruppe Flugpsychologie am Zentrum für Luft- und Raumfahrtmedizin der Luftwaffe aktuell die psychometrischen Merkmale: „Visuelle Gedächtnisfähigkeit“, „Vigilanz“, „Aufmerksamkeitssteuerung“, „Kopfrechnen“, „Räumliches Vorstellungsvermögen“, „Bewegungsantizipationsfähigkeit“, „Informationsverarbeitungskapazität“, „Berufsspezifische Motivation“ und „Leistungsbereitschaft“. Ergänzend sollen nun auch Belastbarkeit, Arbeitshaltung, sowie Kommunikations- und Kooperationsfähigkeit der Bewerber diagnostiziert werden. Mit Hilfe des Zugewinns an Informationen soll eine erhöhte prognostische Validität der Bewerberselektion bewirkt werden. Dies hat zur Folge, dass geeignete Bewerber besser von ungeeigneten differenziert werden können. Es ist zu erwarten, dass dadurch die Drop-out-Raten im späteren Ausbildungsprozess sinken werden.

Um die Fähigkeiten Belastbarkeit, Arbeitshaltung, sowie Kommunikations- und Kooperationsfähigkeit diagnostizieren zu können, soll ein bereits bestehendes Verfahren, der Dyadic Cooperation Test, DCT zum Einsatz kommen. Der DCT wurde im Rahmen einer Projektarbeit des Dezernats Auswahl unter der Leitung von Herrn Mag. Eisl modifiziert, mit dem Ziel einer besseren Anpassung an die Anforderungen der militärischen Flugverkehrskontrolle.

1.2 Der Dyadic Cooperation Test (DCT)

Der DCT ist ein computergestütztes Verfahren der Teamdiagnostik. Die Problemlöseaufgabe bildet die Basis für eine strukturierte Verhaltensbeobachtung zweier Bewerber. Der Test wurde von Dipl.-Psych. Dr. Dirk Stelling im Jahr 1998 entwickelt. Die einzelnen Aufgaben des Tests, wie auch Instruktion und Übungsdurchgänge wurden im Jahr 2013 im Rahmen einer Projektarbeit vom Dezernat Auswahl der Fachgruppe Flugpsychologie am Zentrum für Luft- und Raumfahrtmedizin der Luftwaffe neu modelliert. Das Szenario umfasst ein Verkehrsleitsystem, bei dem vier Straßen mit verschiedenen schweren Fahrzeugen gemäß Vorgabe optimal belastet werden müssen. Dabei kann es zu Fehlermeldungen bei Überlastung

¹ Unter den Flugführungsdiensten (FFD) werden jene Verwendungen zusammengefasst, die die Aufgabe der Überwachung und das unterstützende Eingreifen in Bezug auf die Flugwege von Luftfahrzeugen gemeinsam haben. Dazu zählen u.a. die militärische Flugsicherung, der Einsatzführungsdienst, die Einsatzleitgruppe, der Forward Air Controller wie auch das Bedienpersonal unbemannter Luftfahrzeuge.

einzelner Straßen kommen. Zunächst bekommen die Bewerber eine multimediale Instruktion, im Anschluss daran können noch offene Fragen geklärt werden. Dieser Einführung folgt eine Übungsphase, in der das Verständnis vertieft und die Bedienung der Eingabeelemente gefestigt werden sollen. Danach folgen vier Runden mit Testaufgaben, die zunächst einzeln bearbeitet werden, um so eine Baseline der Bewerberfähigkeit für diesen Aufgabentypus zu gewinnen. Als Vergleichseinheit fungiert der in den Solorunden erspielte Punktwert.

Die eigentliche Aufgabenphase wird zeitgleich von zwei Personen als Teamaufgabe bearbeitet. Hierfür sitzt jeder Bewerber an einem PC-Arbeitsplatz, welcher mit dem benachbarten PC verbunden ist. Einem Probanden ist der Ortsteil City-West, dem anderen City-Ost zugewiesen. Eingaben erfolgen über Nummernpad und die Kommunikation verläuft via Headsets. Nach dem Prinzip des Wechselsprechens kann die Leitung immer nur von einer Seite genutzt werden. In den Teamrunden verhindert eine Trennwand den Sichtkontakt der Probanden. Während der Bearbeitung kommt es immer wieder zu Sondermeldungen, Einschränkungen und möglichen Systemausfällen, auf die es zu reagieren gilt. Außerdem ändern sich auch die Verantwortlichkeiten, also die Rolle des Koordinators beziehungsweise des Zuarbeiters, in jeder Aufgabenrunde. In der Rolle des Koordinators trägt man die Verantwortung für alle Straßen und muss seinen Zuarbeiter dementsprechend anleiten. Als Zuarbeiter nimmt man eine unterstützende Rolle ein. Das Ziel ist es, in allen Durchgängen jede der vier Straßen mit den optimalen Gewichtungen zu belegen. Für die Zielerreichung müssen die Probanden zusammenarbeiten und über die Headsets kommunizieren. Eigenschaften der Kommunikation, Rollenperformanz sowie Einhaltung der gelernten Regeln bilden den Beobachtungsfokus des Verfahrens. Die Probanden werden dabei von einer Auswahlkommission, bestehend aus mindestens einem Psychologen und einem zusätzlichen Experten für die Flugführungsdienste (FFD), dem Prüfstabsoffizier, beobachtet. Im Anschluss werden die gewonnen Eindrücke der Kommission gemeinsam diskutiert und die Bewerber entsprechend der aufgezeichneten Beobachtungen bewertet.

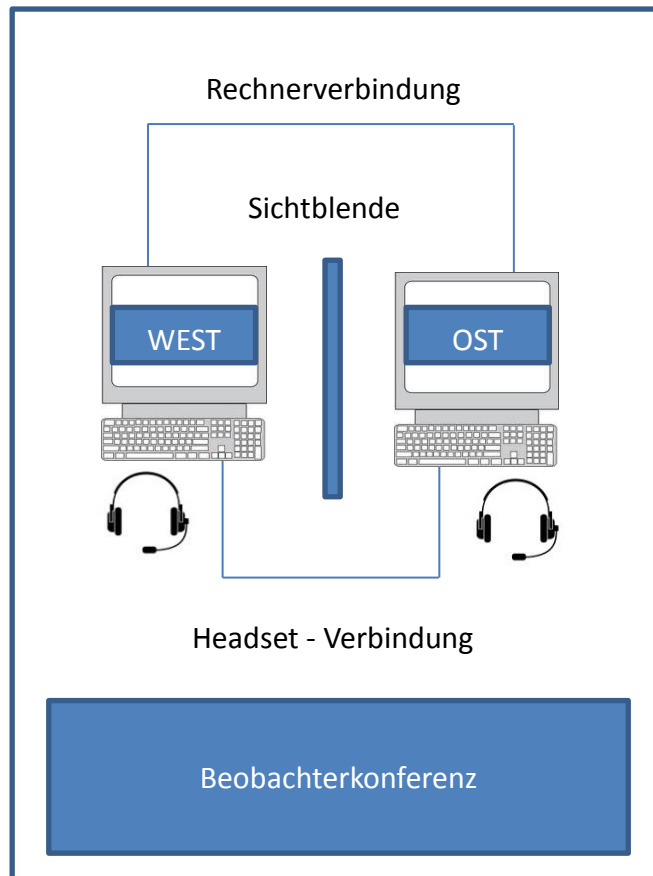


Abbildung 1. Aufbau des DCT

Die Bedingungen des DCT gleichen den realen Tätigkeitsanforderungen der Flugsicherung (Stelling, 1999, S. 24). Die Verhaltensbeobachtung im Kontext des DCT könnte somit auch als eine Sonderform der Arbeitsprobe bezeichnet werden, die laut Jäger (2003, S. 42) in Bezug auf Akzeptanz und das Gütekriterium Validität noch höher als eine herkömmliche Verhaltensbeobachtung anzusiedeln ist. Das DCT-Verfahren in seiner derzeitigen Form verfügt somit über das Potential den Auswahlprozess der FFD zu optimieren. Um eine praktische Einsatzfähigkeit zu erlangen, bedarf es jedoch einer expliziten Operationalisierung der zu erhebenden Konstrukte, sowie einer Schulung aller potenziellen Beobachter im Dezernat Auswahl der Fachgruppe Flugpsychologie am Zentrum für Luft- und Raumfahrtmedizin der Luftwaffe.

1.3 Anforderungsanalysen

Schon während der Projektarbeit wurden zahlreiche Expertengespräche geführt und somit relevante Eigenschaften für die militärischen Flugführungsdienste definiert. Die genannten Fähigkeiten und Fertigkeiten lauten Interaktion, Flexibilität, Arbeitsstil, Entscheidungsfindung und Belastbarkeit. Auch andere Anforderungsanalysen zeigten ähnliche Ergebnisse. Eißfeldt und Heintz veröffentlichten im Jahr 2002 (zitiert nach Hunter & Martinussen, 2008) eine Anforderungsanalyse der Arbeitsplätze von Fluglotsen in der zivilen Flugverkehrskontrolle. Mit Hilfe des Jobanalyse Systems nach Fleishman (1975) konnte gezeigt werden, dass neben den bereits bekannten Merkmalen, wie kognitive Fähigkeiten und Motivation, auch Persönlichkeitseigenschaften von großer Bedeutung sind. Somit sind Kooperation, Kommunikation und der adäquate Umgang mit Stress wichtige Eigenschaften im Anforderungsprofil der Flugverkehrskontrolle. Obwohl sich diese Berufsbezeichnung aus verschiedenen Funktionen zusammensetzt, konnten die Ergebnisse von Eißfeldt und Heintz auf sämtlichen Arbeitsbereiche, also Approach, Aerodrome Control und Area Control, übertragen werden.

1.4 Unfallstatistik

Die Bedeutsamkeit von Kommunikation und Kooperation für die sichere und effiziente Ausführung der Tätigkeit konnte auch in Analysen zur Sicherheitslage der Luftfahrt bestätigt werden. Bei Untersuchungen dieser Art werden die Ursachen von Flugunfällen und Beinahe-Unfällen, so genannten Incidents, analysiert. Beispielsweise befragte die Lufthansa (Hunter & Martinussen, 2008, S. 35) im Zeitraum von 1997 bis 1999 2070 Piloten zu sicherheitskritischen Situationen und konnte verschiedene Gründe für derartige Zwischenfälle aufzeigen. So traten bei 53 Prozent der Incidents Kommunikationsprobleme im Cockpit auf. Hierbei war in den meisten Fällen auch das Flugsicherungspersonal involviert. Laut Selbstausskunft der Piloten wurden wichtige Aussagen oft überhört oder falsch verstanden. Auch mangelnde Kooperation führte nachweislich zu zahlreichen Incidents. In 12 Prozent aller Vorfälle wurde das gemeinsame Handeln nicht genügend koordiniert und entwickelte sich so zum zentralen Problem.

1.5 Zielsetzung der Diplomarbeit

Es wird deutlich, dass auch nicht kognitive Variablen, wie Interaktion- und Kommunikationsfähigkeit, Flexibilität, Arbeitsstil und Belastbarkeit, wichtige Prädiktoren für eine erfolgreiche Tätigkeit in der Flugsicherung sind. Auch im militärischen Bereich sollte die Beurteilung dieser Fähigkeiten in den Auswahlprozess involviert werden. Durch die aktive Unterstützung des DCT-Projekts soll das Verfahren und somit die Erhebung der genannten Konstrukte in der Eignungsdiagnostik der FFD etabliert werden. Eine nachvollziehbare Operationalisierung der psychometrischen Merkmale bildet dabei die Grundlage für eine valide Datenerhebung, deren Objektivität mit Hilfe eines computergestützten Beobachtertrainings erhöht werden soll. Durch den Einsatz des Verfahrens soll eine erhöhte prognostische Validität der Bewerberauswahl erreicht werden. Somit könnte in Zukunft besser zwischen geeigneten und ungeeigneten Bewerbern unterschieden werden, was zur Folge hätte, dass die Drop-out-Raten in der Ausbildung sinken. Dies würde den Ausbildungsprozess ökonomisieren. Diese Daten können jedoch erst nach fortgeschrittener Ausbildung der Bewerber erhoben werden und sind somit kein Gegenstand dieser Studie. Eine dafür erforderliche Analyse folgt zu einem späteren Zeitpunkt mit den Daten aus der Langzeiterfolgskontrolle (LEK).

1.6 Gliederung der Diplomarbeit

Die Diplomarbeit teilt sich in sieben Kapitel. Nach der Einleitung werden in Kapitel zwei die theoretischen Grundlagen der Thematik genauer beleuchtet. Das Szenario des DCT umfasst eine dynamische Teamaufgabe, welche nur durch effektive Kommunikation gelöst werden kann. Dabei wird das Verhalten der Bewerber von einer geschulten Beobachterkommission erfasst und anschließend bewertet. Im Fokus des Theorieteils stehen somit Interaktion und Kommunikation, die wissenschaftliche Verhaltensbeobachtung, die theoretischen Hintergründe der zu erfassenden Dimensionen, Belastbarkeit und Arbeitsstil sowie die Grundlagen eines Beobachtertrainings. Die Hauptinhalte des dritten Abschnitts dieser Diplomarbeit bilden die aus dem aktuellen Forschungsstand resultierenden Fragestellungen. Spezifiziert werden die Leitfragen durch die entsprechenden Hypothesen. Die Methoden zur Prüfung dieser Hypothesen werden im vierten Kapitel angeführt und

die Ergebnisse im darauf folgenden Abschnitt, Kapitel fünf, präsentiert. Das sechste Kapitel beinhaltet die Diskussion der Diplomarbeit. Hier werden die Resultate bewertet und in Beziehung zum bisherigen Forschungsstand gesetzt, um anschließend deren zukünftige Bedeutung zu akzentuieren. Abschließend erfolgt eine Zusammenfassung der Arbeit. In diesem Abschnitt werden die wichtigsten Punkte resümiert und ein Ausblick auf mögliche Fragen zukünftiger Arbeiten, welche sich aus den neuen Erkenntnissen ergeben könnten, aufgezeigt. Im Anhang dieser Diplomarbeit befinden sich die gelisteten Indikatoren der konstruierten Skalen, das Quellenverzeichnis und die eidesstaatliche Erklärung bezüglich der selbstständigen Erarbeitung der vorliegenden Inhalte.

2. THEORIE

Das zweite Kapitel der Diplomarbeit bildet den Theorieteil. Dieser setzt sich mit dem bisherigen Forschungsstand auseinander, zeigt wissenschaftliche Grundlagen und bisherige Erkenntnisse bezüglich der im Interessensfokus stehenden Dimensionen Belastbarkeit, Arbeitsstil und Interaktion. Des Weiteren thematisiert dieses Kapitel die methodischen Grundlagen der wissenschaftlichen Verhaltensbeobachtung sowie die theoretischen Grundlagen zur Konstruktion eines geeigneten Beobachtertrainings.

2.1 Dimensionen

Zunächst muss geklärt werden, welche psychologischen Konstrukte mit Hilfe der Verhaltensbeobachtung im Rahmen des DCT erfasst werden sollen. Hierzu wurde im Rahmen von Projektarbeiten im Dezernat Auswahl des ZentrLuRMedLw bereits viel Vorarbeit geleistet. Aus Experteninterviews im Rahmen einer gerade laufenden Anforderungsanalyse der militärischen Flugverkehrskontrolle und des Einsatzführungsdienstes konnten bereits bedeutsame Eigenschaften für die Tätigkeiten der Flugführungsdienste abgeleitet werden.

Interaktion, Flexibilität, Arbeitsstil, Kommunikationsfähigkeit und Belastbarkeit sind latente psychologische Konstrukte, die nicht auf direktem Wege zugänglich sind. Im

ersten Schritt müssen diese Verhaltensdimensionen definiert werden. Eine gründliche Bedeutungsanalyse klärt die Begrifflichkeit und bildet die Grundlage für eine analytische Definition. Die Angabe der Operationen, die zur Erfassung der durch den Begriff bezeichneten Sachverhalte notwendig sind, bezeichnet man als eine operationale Definition (Bortz & Döring, 1995, S.63). Will man theoretischen Konstrukte messbar machen, müssen entsprechende Indikatoren gefunden werden. Diese werden aus typischen Verhaltensweisen abgeleitet. Solche können sowohl verbaler als auch nonverbaler Natur sein (Becker-Beck, 1997, S.117). Durch eine anschließende Strukturierung und Klassifizierung können die Verhaltensanker den jeweiligen Konstrukten zugeteilt werden. Beim Operationalisieren ist es zudem notwendig, eine passende Erhebungsmethode und entsprechende Messinstrumente genau zu definieren.

2.1.1 Belastungsfähigkeit

Eine Belastung sind physische und psychische Faktoren, die auf den Menschen einwirken. Sie treten unter belastenden Bedingungen, wie zum Beispiel Zeitdruck, Leistungsdruck oder Mehrfachanforderung, auf (Schrott, 2003, S. 56 f)

„Belastbarkeit ist das Ausmaß, in dem es einer Person gelingt, Belastungen standzuhalten, und ihren Einfluss auf Leistungsfähigkeit und Befindlichkeit so gering wie möglich zu halten.“ (Schrott, 2003, S.56)

In einer schematischen Darstellung analysiert Wieland-Eckelmann (1997, S.35) das Konstrukt Belastungsfähigkeit auf den Dimensionen Belastungsanalyse, Beanspruchungsanalyse sowie Bewertungs- und Folgeanalyse. Die Belastungsanalyse bezieht sich dabei auf die gegebenen Anforderungen und die individuellen Leistungsvoraussetzungen. Das Zeit und Ressourcenmanagement, also die Inanspruchnahme der Leistungsvoraussetzungen, wird im Kontext von Beanspruchungsanalysen genauer betrachtet. Für die Verhaltensbeobachtung sind Bewertungs- und Folgeanalyse von zentraler Bedeutung, da sich diese mit den beobachtbaren Folgen von Belastungen auseinandersetzen. Die Autoren unterscheiden hierbei die Komponenten physiologische Reaktion, subjektive Befindlichkeit und Leistungsfähigkeit.

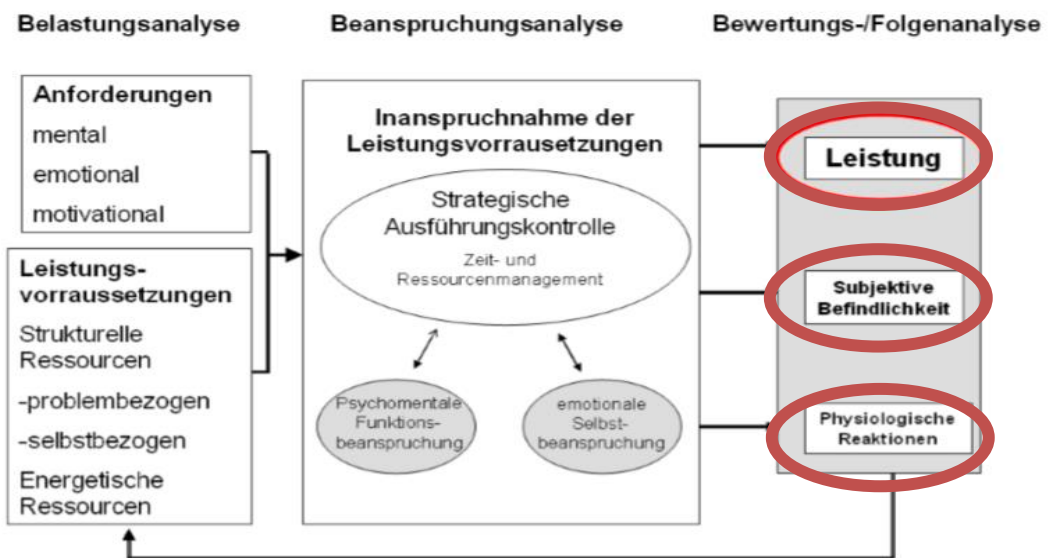


Abbildung 2. Unterscheidung Belastungsanalyse, Beanspruchungsanalyse und Bewertungs-/Folgenanalyse (Wieland-Eckelmann, 1992, S.35)

2.1.1.1 Physiologische Reaktion

In frühen Phasen unserer Spezies waren Stressoren meist direkt mit der Bedrohung der leiblichen Unversehrtheit verbunden. Evolutionsbedingt kommt es in Situationen, welche von einer erhöhten Belastung gekennzeichnet sind, bei allen Menschen zu ähnlichen physiologischen Reaktionen. Diese stellen eine adaptive Vorbereitung auf körperliche Angriffe dar und äußern sich auch in der sogenannten Kampf-oder-Flucht-Reaktion. Obwohl sich die Art der Belastungen im Laufe der Menschheitsgeschichte stark verändert hat, sind diese Muster heute noch deutlich nachweisbar. Das neuronale Zentrum der Stressreaktion bildet der Hypothalamus, welcher ebenfalls an einer Vielzahl von emotionalen Reaktionen beteiligt ist. Durch die stressbedingte Aktivierung können zahlreiche physiologische Stressreaktionen beobachtet werden. Diese Prozesse betreffen vor allem das autonome Nervensystem und lauten unter anderem: erhöhte Schweißabsonderung, Gänsehaut, erhöhte Herzraten, erweiterte Pupillen, schnellere und tiefere Atmung, Entspannung der Blase, Zittern sowie feuchte Hände (Gerrig & Zimbardo, 2008, S.470 und Schwarzer, 1981, S. 81).

2.1.1.2 Subjektive Befindlichkeit

Die subjektive Befindlichkeit kann in die beiden Komponenten Emotion und Kognition unterteilt werden. Zu den emotionalen Indikatoren von Beanspruchung zählen: Gefühle der inneren Unruhe, Nervosität, Gefühle des Gehetztseins, Hilflosigkeit, Unzufriedenheit, Ärger, Ängste, Selbstvorwürfe, Gedankenschleifen, Denkblockaden und Blackouts. Mit diesen emotionalen Beanspruchungen gehen meistens auch kognitive Einschränkungen einher. So kann es ab einer bestimmten Intensität der Belastung zu Wahrnehmungs-, Koordinations- und Konzentrationsstörungen kommen (Gros, 1994, S.115).

2.1.1.3 Leistungsfähigkeit

Handlungssteuernde Emotionen, wie ein starkes Angstempfinden und kognitive Einschränkungen, können sich auf die Leistungsfähigkeit einer Person auswirken. Beanspruchungsreaktionen sind individuell. Ab einer bestimmten Ausprägung können diese Reaktionen einen negativen Einfluss auf die Leistungsfähigkeit einer Person haben. Diese Annahme wird mit Hilfe des Yerkes-Dodson-Gesetzes untermauert werden. Im Speziellen beschreibt die Theorie von Yerkes und Dodson aus dem Jahre 1908 die kognitive Leistungsfähigkeit einer Person in Abhängigkeit von unterschiedlichen Erregungsniveaus. Der Zusammenhang zwischen physiologischer Aktivierung und der Leistungsfähigkeit ist demnach umgekehrt U-förmig. Somit ist der Leistungsverlauf je nach Grad der aktuellen Aktivierung dynamisch. Bei Unterforderung unterliegt der Mensch einem Erregungsniveau niedriger Ausprägung. Es entsteht ein Leistungstiefstand. Einer Steigerung der emotionalen Erregung folgt wiederum eine entsprechende Erhöhung der Leistungsfähigkeit. Übersteigt diese jedoch das individuell variierende Optimum, sinkt die Leistungsfähigkeit wieder ab (Yerkes & Dodson, 1908).



Abbildung 3. Leistungsfähigkeit und Belastung (Litzcke & Schuh, 2010, S.12)

Will man Belastungsfähigkeit beobachten, gibt es also verschiedene Indikatoren, welche auf den vier Messebenen sprachliche Mitteilung, körperliche Erregung, offenes Verhalten und Ausdruck erfasst werden können. Auch die nonverbalen Informationen, wie Körperhaltung (gebeugt, starr), die Stimme (leise, stotternd, desorganisiert) und der Gesichtsausdruck (aufgerissene Augen), bilden in diesem Zusammenhang eine wichtige Beobachtungsgrundlage, von der man bis zu einem bestimmten Grad auf die inneren Angstzustände einer Person schließen kann (Schwarzer, 1981, S.81).

2.1.2 Arbeitsstil

Unter Arbeitsstil versteht man die Art und Weise, wie eine Person mit Leistungsanforderungen umgeht (Wagner-Menghin, 2011, S.16). Neben kognitiven sind auch nicht-kognitive Variablen zur Erklärung von Leistungsverhalten von zentraler Bedeutung. Diese Annahme spiegelt sich auch im Wiener Diagnosemodell von Hochleistungspotential wieder. Das differentielle Modell zur Diagnose von Hochbegabung schreibt neben der intellektuellen Begabung auch nicht intellektuellen Fähigkeiten, wie der Anerkennung und Entwicklung der eigenen Leistungsfähigkeit, einen hohen Stellenwert zu. Bei den nichtkognitiven Fähigkeiten handelt es sich gemäß dem Wiener Diagnosemodell um Persönlichkeit sowie die soziokulturelle Anregungsumwelt im Kontext der Leistungserbringung (Holocher-

Ertl, 2008). Die unbestreitbare Relevanz des Arbeitsstils in Leistungssituationen konnte zudem durch empirische Untersuchungen bestätigt werden. Unter anderem gelang dies Lounsbury, Sundstorm, Loveland und Gibson in einer Studie aus dem Jahr 2003. Bezüglich der erreichten Abschlussnote in einem Hochschulkurs konnten 16 % der Varianz durch kognitive und weitere 7 % der Varianz durch die Variablen des Fünf-Faktoren-Modells erklärt werden.

Die fünf Faktoren beziehen sich auf das anerkannte Big-Five-Persönlichkeitsmodell, welches unter anderem den Arbeiten von Norman aus dem Jahr 1963 entspringt. Norman verfolgte den Ansatz der Sedimentationshypothese, wonach alle relevanten Unterschiede zwischen Individuen folglich auch in die menschliche Sprache eingegliedert wurden. Laut diesem lexikalischen Ansatz sollten verschiedene Persönlichkeitsmerkmale je nach Bedeutsamkeit durch eine unterschiedliche Anzahl an Wörtern und Synonymen in der Sprache präsent sein (Bartussek & Amelang, 1981, S. 346). McCrae & Costa griffen diese früheren Arbeiten auf und untersuchten verschiedene Wortlisten faktorenanalytisch. Mit Hilfe dieser Untersuchungen konnte gezeigt werden, dass sich die charakterlichen Eigenschaften eines Menschen im Wesentlichen durch die Ausprägung der Faktoren Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit unterscheiden (Borkenau, P. & Ostendorf, F., 1993). Man geht davon aus, dass die Persönlichkeit einer Person anhand dieser fünf Faktoren und den entsprechenden Facetten (siehe Abbildung 2) hinreichend beschrieben werden kann.

Für die Erklärung von Leistungsverhalten werden den Facetten Leistungsstreben, Selbstdisziplin, Pflichtbewusstsein des Faktors Gewissenhaftigkeit sowie Impulsivität und Ängstlichkeit des Faktors Neurotizismus eine besondere Bedeutung zugeschrieben (Wagner-Menghin, 2011, S.16).

Tabelle 1. Facetten des Fünf -Faktoren-Modells nach Costa & McCrae (1992, zitiert nach Maltby et al., 2011). Facetten mit größten Erklärungsgehalt für Leistungsstreben hervorgehoben.

Neurotizismus	Extraversion	Offenheit für Erfahrung	Verträglichkeit	Gewissenhaftigkeit
Ängstlichkeit	Herzlichkeit	Offenheit für Fantasie	Vertrauen	Kompetenz
Reizbarkeit	Geselligkeit	Offenheit für Ästhetik	Freimütigkeit	Ordnungsliebe
Depression	Durchsetzungs-fähigkeit	Offenheit für Gefühle	Altruismus	Pflichtbewusstsein
Soziale Befangenheit	Aktivität	Offenheit für Handlungen	Entgegenkommen	Leistungsstreben
Impulsivität	Erlebnishunger	Offenheit für Ideen	Bescheidenheit	Selbstdisziplin
Verletzlichkeit	Frohsinn	Offenheit für Werte- und Normsysteme	Gutherzigkeit	Besonnenheit

Die Ausführungen zeigen deutlich, Theorie und Praxis konnten die Bedeutung nicht kognitiver Fähigkeiten im Kontext von Leistungssituationen aufzeigen. Dennoch fehlt es bis heute an einer exakten und einheitlichen Definition des Konstrukts Arbeitsstil. Häufig werden neben den angeführten fünf Faktoren der Persönlichkeit motivationspsychologische Inhalte und kognitive Stile mit dem Begriff Arbeitsstil in Verbindung gebracht. So auch von Kubinger und Ebenhöf (1997), welche die objektive Persönlichkeitsbatterie Arbeitshaltungen, in der Kurzform AHA genannt, konstruierten. Die darin enthaltenen Subtests sollen die motivationspsychologischen Konstrukte Leistungsmotivation, Frustrationstoleranz und Anspruchsniveau sowie den kognitiven Stil Impulsivität vs. Reflexivität erfassen.

2.1.2.1 Motivationspsychologischer Ansatz

Die angeführten Konstrukte sind gemäß dem motivationspsychologischen Ansatz voneinander abhängige Gebilde, welche sich auf höchst komplexer Art und Weise gegenseitig beeinflussen. Leistungsmotiv, Anspruchsniveau und Frustrationstoleranz stehen in ständiger Wechselwirkung und sollten somit stets einer gemeinsamen Gesamtbetrachtung unterzogen werden (Harbourn, 2007, S.14).

Leistungsmotivation bzw. Leistungsmotiv

Im Gegensatz zur Motivation, also der momentanen Tendenz zu einer ganz bestimmten Handlung, beschreibt das Motiv situationsüberdauernde Antriebe, welche das Verhalten einer Person über einen längeren Zeitraum beeinflussen. Ungeachtet der inhaltlichen Abgrenzung, wird in der Literatur häufig der Begriff Leistungsmotivation gebraucht. Trotz der abweichenden Begrifflichkeit geht man bei Leistungsmotivation von einem Motiv, also von einem über die Zeit relativ stabilen Konstrukt, aus. Die Bezeichnung Leistungsmotiv ist in diesem Zusammenhang wohl treffender. Im Jahr 1938 konnte Henry Murray zeigen, dass sich dieses Motiv von Mensch zu Mensch unterscheidet und ganz unterschiedliche Neigungen zur Folge hat (Gerrig & Zimbardo, 2008, S. 442). Eckhardt und Schuler (1995) definierten den Begriff wie folgt: *„Aspekte der Leistungsmotivation sind das selbstgesetzte Anspruchsniveau (z. B. bezüglich der Menge der angestrebten Leistung, bezüglich der Genauigkeit und Gründlichkeit der angestrebten Leistung) und die Beharrlichkeit der Zielverfolgung angesichts äußerer und innerer Widerstände“* (zitiert nach Ebenhöf & Kubinger, 1997). Zentrale Punkte der Leistungsmotivation sind demnach das Ausmaß der persönlichen Zielsetzung und die individuelle Ausdauer, mit welcher diese Ziele verfolgt werden.

Anspruchsniveau

Das Anspruchsniveau wird als der *„für ein Individuum charakteristische Gütegrad, bezogen auf die erreichte Leistungsfähigkeit, der für die Selbstbewertung eines erzielten Handlungsresultates entscheidend ist“*, beschrieben (Heckhausen 1955, S. 119).

Ausschlaggebend für die Ausprägung des Anspruchsniveaus sind individuelle Erfahrungen. Mit dem Erleben von Erfolg steigt das Anspruchsniveau, parallel dazu sinkt es durch Misserfolg gleichermaßen ab. Erfolg bedeutet, dass eine Person ihr zuvor definiertes Anspruchsniveau erreichen kann (Harbourn, 2007, S.12). Das Anspruchsniveau beschreibt die Einstellung zu Erfolg beziehungsweise zu Misserfolg sowie dem entsprechenden Schwierigkeitsgrad einer Aufgabe, welche von Personen individuell präferiert wird. Die beiden zentralen Leistungsmotive setzen sich aus „Hoffnung auf Erfolg“ und „Furcht vor Misserfolg“ zusammen. Beim erstgenannten Motiv handelt es sich um sogenannte erfolgssuchende Personen. Aufgaben, welche der eigenen Leistungsfähigkeit entsprechen, werden von

Erfolgssuchenden gezielt ausgewählt, um sich in diesen zu bewähren. Hingegen werden fähigkeitsentsprechende Aufgaben von Personen mit dem Motiv „Furcht vor Misserfolg“, welche auch als „Misserfolgsvermeider“ bezeichnet werden, ausdrücklich gemieden. Sie suchen Anforderungen deutlich unterhalb des eigenen Leistungsniveaus, um ein mögliches Scheitern zu vermeiden (Kubinger, 2009, S. 239).

Frustrationstoleranz

Frustration entsteht in Situationen, in welchen Menschen in ihrer individuellen Zielerreichung, der eigenen Motivbefriedigung, behindert oder von ihr abgebracht werden (Dollard, 1939, S. 7). Die Frustrationstoleranz beschreibt den Umgang mit Situationen dieser Art und wird definiert als *„Eigenschaft einer Person, Frustrationen über einen längeren Zeitraum auszuhalten, ohne das Ziel der Motivbefriedigung, der Motivation, aufzugeben“* (Karner & Wagner, 2007, S. 6.). Kann eine Person trotz Frustration ihre Leistungsmotivation aufrechterhalten, verfügt sie demnach über eine hohe Frustrationstoleranz. Kommt es in Folge von frustrierenden Situationen jedoch zur Minderung oder gar dem Verlust der Leistungsmotivation, ist die Toleranzfähigkeit gegenüber Frustrationen von geringer Ausprägung. Im Kontext des motivationspsychologischen Ansatzes kann der Begriff Frustration auch auf das Nichterreichen des individuell gesetzten Anspruchsniveaus übertragen werden.

2.1.2.2 Kognitive Stile

Kognitive Stile gelten als formale Besonderheiten der Informationsverarbeitung. Sie beschreiben die Art und Weise, beziehungsweise den Denkstil, in dem die individuelle Verarbeitung einer Person verläuft. Dabei kann nicht generell, sondern nur in Bezug auf einen ganz bestimmten Aufgabentyp zwischen besseren oder schlechteren Denkstilen unterschieden werden (Kubinger, 2009, S. 250-252). Zu den bedeutendsten kognitiven Stilen in Bezug auf das Konstrukt Arbeitsstil zählen die Feldabhängigkeit und die Typisierung impulsiv vs. reflektierend. Sie sind aus diesem Grund in Tabelle 2 farblich hervorgehoben.

Tabelle 2. Kognitive Stile (Häcker & Stampf, 2009, S. 202 ff)

Kognitiver Stil	Prägende ForscherInnen
Analytisch	
Feldunabhängigkeit vs. global	
Feldabhängigkeit	Jaensch, Selinka, Witkin (1954)
impulsiv vs. reflektierend	Kagan (1965)
Interferenzeignung	Hörmann (1960)
nivellieren vs. akzentuieren	
Toleranz vs. Intoleranz gegenüber Instabilität	
Physiognomisches vs. Realistisches Erfassen	
verdrängen vs. sensitiveren	Byrne (1961)

Feldabhängigkeit

Feldabhängigkeit ist das Ausmaß der Ablenkungsfähigkeit durch die Wahrnehmungsumgebung, also das Feld während der Aufgabenbearbeitung. Eine feldabhängige Person lässt sich schnell von der äußeren Umgebung ablenken und verliert somit das wahrzunehmende Ziel aus den Augen. Feldunabhängige Personen sind im Gegensatz dazu ablenkungsresistent und können sich vollkommen auf das Ziel konzentrieren, indem sie die Wahrnehmungsumgebung zum größten Teil ausblenden (Kubinger, 2009, S.251). Goodenough (1978) differenziert Personen nach der Verwendung unterschiedlich lokalisierter Bezugssysteme. Demnach verfügen feldunabhängige Personen über internale frames of reference, womit sie eingehende Informationen verarbeiten. Feldabhängige Personen nutzen hingegen external frames of references. Sie orientieren sich weniger aktiv an der Verarbeitung eingehender Informationen. Kritiker bemängeln am kognitiven Stil der Feldabhängigkeit die unklare Abgrenzbarkeit zu bereits bekannten Persönlichkeitskonstrukten, wie zum Beispiel der Intelligenz (Adlmann & Gittler, 2001, S.170).

Impulsivität versus Reflexivität

Der Denkstil Impulsivität versus Reflexivität beschreibt Schnelligkeit sowie Genauigkeit der Aufgabenbearbeitung. Diese kann entweder schnell und fehlerreich oder langsam und fehlerarm sein. Auch hier gilt, keine der beiden Stile ist generell besser oder schlechter. Nur in Bezug auf bestimmte Aufgabentypen ist eine impulsive, beziehungsweise reflexive Arbeitsweise von Vorteil (Kubinger, 2009, S. 252).

2.1.2.3 Flexibilität

Flexibel stammt aus dem Lateinischen und kann mit den deutschen Wörtern biegsam, anpassungsfähig übersetzt werden. Schmuck (1996, S. 11) unterscheidet zwei verschiedene Arten von Flexibilität. Eine wird als herausgeforderte Verhaltensoptimierung definiert. Diese Art der Flexibilität beinhaltet die Anpassung an sich veränderte Umweltbedingungen zur Zielerreichung. Die andere Art der Flexibilität beinhaltet die Fähigkeit der spontanen Verhaltensoptimierung ohne das unmittelbare Vorhandensein eines Selektionsdrucks. In Bezug auf das Konstrukt Arbeitsstil steht diese Art der Spontanreflexivität im Betrachtungsfokus. *„Flexibles Verhalten drückt sich darin aus, daß [sic!] bei einer Anforderung mit gegebener Bearbeitungsroutine diese Routine spontan, d.h. ohne Aufforderung und ohne objektive Notwendigkeit verlassen wird zugunsten eines neuen Bearbeitungswegs, welcher ein effizienteres Vorgehen ermöglicht als der gegebene“* (Schmuck, 1996, S.11). Flexibilität zählt somit zu den nichtkognitiven Fähigkeiten einer Person, welche durch Strategieanpassung Leistungsverhalten beeinflussen kann. Sie kann somit ebenfalls dem Konstrukt Arbeitsstil zugeordnet werden.

2.1.3 Interaktion & Kommunikation

Kommunikation ist *„was jemand von sich gibt bzw. das, was beim anderen ankommt“* (Schulz von Thun, 1992, S. 9). Zwischenmenschlicher Austausch verläuft also über die Vermittlung von Nachrichten. Dabei differenziert man die Makro-Ebene, also Kommunikation und Sprache, von der Mikro-Ebene der einzelnen Nachricht. Im entsprechenden Kommunikationsmodell unterscheidet Schulz von Thun vier Seiten einer Nachricht. Sachinhalt, Selbstoffenbarung, Beziehung und Appell (siehe Abbildung 6). Der Sachinhalt ist der Informationsteil einer Nachricht.

Mit Hilfe der Selbstoffenbarung gibt der Sender etwas von sich selbst preis. Die Beziehung verdeutlicht, wie Sender und Empfänger zueinander stehen. Die vierte Seite einer Nachricht bildet laut Schulz von Thun der Appell. Er beschreibt, wozu der Sender den Empfänger bewegen möchte.

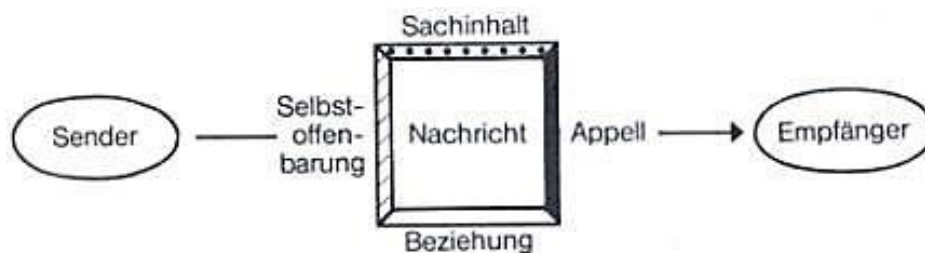


Abbildung 4. Vier Seiten einer Nachricht (Schulz von Thun, 1992)

Die Güte einer Kommunikation resultiert demnach auch aus der Qualität der Kommunikationskette. Diese setzt sich aus Verschlüsselung der Nachricht, Versendung und Empfang sowie anschließender Entschlüsselung und Interpretation des Empfängers zusammen. Die wichtigsten Fähigkeiten, um einen reibungslosen Ablauf der Kommunikationskette nach Schulz von Thun zu gewährleisten, bilden mündlicher Ausdruck, Klarheit der Sprache sowie die Spracherkennung. Die Fähigkeit, sich mündlich auszudrücken, beinhaltet Wörter und Sätze so zu verwenden, dass sie der Kommunikationspartner versteht. Die Klarheit der Sprache bezieht sich auf die Ausdrucksweise, die Satzkonstruktion und den Einsatz bedeutungsäquivalenter Wörter. Für den Empfang der Nachricht benötigt man die Fähigkeit der Spracherkennung. Diese beinhaltet, die Nachricht einer anderen Person wahrzunehmen und diese zu verstehen, also die versendete Nachricht korrekt zu entschlüsseln. Hierfür ist auch ein gewisses Maß an kognitiver Verarbeitung notwendig (Fleischmann, 2010, S. 55-56).

Die Interdependenz, also die Abhängigkeit der Interaktionspartner, ist ein wesentliches Merkmal in sozialen Interaktionen einer Gruppe. Aus diesem Grund lassen sich die Verhaltensweisen der Gruppenmitglieder nicht unabhängig voneinander betrachten und müssen zueinander in Beziehung gesetzt werden. Die

konzeptuelle Abgrenzung sozialer Interaktion von Kommunikation ist fließend, weshalb die beiden Begriffe auch häufig synonym verwendet werden. Eine soziale Interaktion ist immer auch eine Form der Kommunikation, also ein Austausch von Informationen (Becker-Beck, 1997, S. 22).

Auf Grundlage der gegebenen Gruppenstruktur teilt Schelling (1960) die Ausprägung der gegenseitigen Abhängigkeit in 3 Möglichkeiten:

<i>Kooperative Interdependenz</i>	(Des einen Gewinn ist auch des anderen Gewinn.)
<i>Konkurrierende Interdependenz</i>	(Des einen Gewinn ist des anderen Schaden.)
<i>Zielambivalente Interdependenz</i>	(Des einen kleiner Gewinn ist des anderen kleiner Gewinn, des einen großer Gewinn ist des anderen großer Schaden.)

Zusätzliche kann es zu prozessualen Interdependenzen kommen, wenn das individuell wahrgenommene Verhalten des Interaktionspartners das eigene Verhalten beeinflusst. Auch die Gruppeninteraktion kann in Hinblick auf dieser Ebene betrachtet werden. Der Prozessaspekt erfasst die dynamischen, ständig variierenden Systemmerkmale einer Gruppe und ihre Interaktionsmuster. Je nach Größe der Analyseeinheit können unterschiedliche Phänomene der Interaktion beobachtet werden. Man unterscheidet Dyaden, Triaden, Tetraden und folgende. Einen Überblick über mögliche Interaktionsmuster in Dyaden gibt Becker-Beck (1997, S.29). Sie unterscheidet:

1. *Reziprozität und Symmetrie versus Kompensation und Komplementarität*
2. *Kooperation, Wettbewerb und Konflikt*
3. *Kontrolle und Dominanz*
4. *Interpersoneller Einfluss und Konformität*

2.1.3.1 Reziprozität & Symmetrie vs. Kompensation & Komplementarität

In dyadischen Interaktionen kann das Auftreten eines Verhaltensaktes einen förderlichen oder einen hemmenden Effekt auf die Ausprägung des folgenden

Verhaltens haben. So führt ein freundliches Lächeln mit einer hohen Wahrscheinlichkeit zu einer freundlichen Reaktion, wobei Aggressivität eher einen hemmenden Einfluss auf anschließende Verhaltensweisen aufweist. Man spricht von Reziprozität, wenn Interaktionspartner qualitativ ähnliche Verhaltensweisen austauschen. In Abgrenzung dazu bezieht sich der Begriff Symmetrie auf die quantitative Ausgeglichenheit der Interaktionsbeiträge. Dem gegenüber steht die komplementäre Interaktion. Hierbei zeigen sich die Personen gegenseitig ein qualitativ abweichendes Verhalten. Kompensation bezeichnet im speziellen die Hemmung eines gleichartigen Verhaltens.

2.1.3.2 Kontrolle & Dominanz

Soziale Kontrolle ist die „*Gesamtheit der Zwänge..., die Menschen durch das, was sie sagen, und dadurch, wie sie ihre Kommunikation strukturieren, aufeinander ausüben.*“ (Wiemann & Giles, 1990, S. 214)

Besitzt eine Person mehr soziale Kontrolle als die andere Person, bezeichnet man diesen Zustand als Dominanz. Wird dominantes Verhalten gezeigt, steigt die Auftrittswahrscheinlichkeit einer bestimmten Verhaltensweise des Partners. Dabei kann Dominanz als stabile Persönlichkeitseigenschaft oder als temporäres Interaktionsverhalten betrachtet werden. Dem gegenüber steht das Konstrukt der Submissivität. Eine submissive Person verfügt über weniger soziale Kontrolle und fügt sich der dominanteren Person (Becker-Beck, 1997, S. 40-42).

2.1.3.3 Kooperation, Wettbewerb & Konflikt

Kooperation und Wettbewerb sind Situationsmerkmale, die sich durch die Konstellation der jeweiligen Ziele der Interaktionspartner unterscheiden (Deutsch, 1949). In kooperativen Situationen kann ein Gruppenmitglied seine eigenen Ziele nur erreichen, wenn die anderen Mitglieder dieses Ziel ebenfalls erreichen. Es besteht eine gleichgerichtete Wechselbeziehung. In Wettbewerbssituationen sind die Ziele der Gruppenmitglieder gegenläufig. Jedes Mitglied kann sein Ziel nur erreichen, wenn die anderen ihre Ziele nicht erreichen. Dabei müssen die Ziele der Mitglieder nicht unbedingt identisch sein. Für einen Beobachter sind diese Verhaltensintentionen kaum zugänglich. Man kann auf kooperative Interaktionsmuster

schließen, wenn in aufgabenbezogenen Akten Reziprozität oder Komplementarität zu beobachten ist. Ein Beispiel für komplementäre Interakte wäre, wenn aufgabenbezogene dominant-freundliche Verhaltensweisen durch submissiv-freundliche Akte ergänzt werden. Im Gegensatz dazu können in Wettbewerbssituationen vermehrte Reziprozität bei Streitakten und mangelnde Unterstützung bei aufgabenbezogenen Inhalten erwartet werden. Im Unterschied zum Wettbewerb liegen bei Konflikten nicht inkompatible Ziele, sondern inkompatible Handlungstendenzen oder antagonistische Interaktionen vor. Auf der Inhaltsebene wird ein Konflikt aufgrund von unterschiedlichen Meinungen, Einstellungen oder Werthaltungen ausgelöst. Dies äußert sich auf der Verhaltensebene zum Beispiel, wenn beide Personen Dominanz beanspruchen oder Streitakte senden. Dennoch ist zu beachten, dass Konflikte sowohl in Kooperations- und in Wettbewerbssituationen entstehen können, obwohl man dennoch davon ausgehen kann, dass der Konfliktbegriff der Wettbewerbssituation deutlich näher steht (Deutsch, 1976, S.18).

2.1.3.4 Interpersoneller Einfluss & Konformität

„Unter Konformität wird die Veränderung individueller Verhaltensweisen, Überzeugungen, Einstellungen etc. infolge von sozialer Beeinflussung (...) verstanden.“ (Stiem & Stürmer, 2013, S. 24)

Konformität beschreibt also die Änderung des Verhaltens in Richtung der Verhaltensweisen der anderen Gruppenmitglieder. Dieser Änderungsprozess kann nur beobachtet werden, wenn das Ausgangsverhalten bekannt ist. Der interpersonelle Einfluss, der zu solchen Änderungsprozessen führt, wird in die vier Typen Uniformität, Antikonformität, Unabhängigkeit und Konformität unterteilt (Hollander & Willis, 1967).

2.2 Verhaltensbeobachtung

Das Beobachten von Verhalten ist ein Wahrnehmungsprozess, bei dem Menschen ihre Umwelt ständig erfassen (Faßnacht, 1995, S. 13). Eine Verhaltensbeobachtung ist also ein eher unspezifischer Begriff, der in der Literatur unterschiedlich definiert wird. In einer dieser Definitionsmöglichkeiten von Graumann (1966) wird eine Verhaltensbeobachtung wie folgt umschrieben:

„Die absichtliche, aufmerksam-selektive Art des Wahrnehmens, die ganz bestimmte Aspekte auf Kosten der Bestimmtheit von anderen beachtet, nennen wir Beobachtung. Gegenüber dem üblichen Wahrnehmen ist das beobachtende Verhalten planvoller, selektiver, von einer Suchhaltung bestimmt und von vorneherein auf die Möglichkeit der Auswertung des Beobachteten im Sinne der übergreifenden Absicht gerichtet. Im alltäglichen Verhalten gehen Wahrnehmen und Beobachten fast unmerklich ineinander über.“ (Graumann, 1966, S. 86)

Die zentralen Punkte dieser Definition einer Verhaltensbeobachtung setzen sich aus Absicht, Selektion und Auswertung zusammen. Die Absicht beschreibt die Tatsache eines geplanten Vorgehens, welchem ein bestimmter Zweck und ein entsprechendes Ziel vorausgesetzt werden. Dabei werden nur einige Aspekte der Gesamtheit aller wahrzunehmenden Inhalte genauer betrachtet und unterliegen somit einem Selektionsprozess. Die Auswertung bildet den dritten Punkt in der Begriffsbestimmung von Graumann. Im Detail besagt Auswertung, dass die aus einer Verhaltensbeobachtung resultierenden Ergebnisse immer auf eine mögliche Auswertbarkeit ausgerichtet sind. Die wahrgenommenen Inhalte müssen also mit Hilfe von entsprechenden Zeichensystemen, die eine bestimmte Bedeutung tragen, abgebildet werden.

Dieses alltägliche Vorgehen geschieht meist unstrukturiert und muss vom Begriff der wissenschaftlichen Verhaltensbeobachtung abgegrenzt werden. Eine spontane Zufallsbeobachtung wird als Gelegenheitsbeobachtung bezeichnet. In Abgrenzung dazu ist die strukturierte Beobachtung ein geplanter Wahrnehmungsprozess (Hartmann, 1970, S. 197). Um das Kriterium der Wissenschaftlichkeit zu erfüllen, muss eine Beobachtung das Ziel haben, eine wissenschaftliche Theorie, beziehungsweise daraus abgeleitete Hypothesen zu prüfen und dabei eine replizierbare sowie objektive Durchführung gewährleisten. Eine wissenschaftliche Beobachtung ist dementsprechend immer auch eine systematische Beobachtung und eine grundlegende Methode der Psychologie (Greve & Wentura, 1997, S. 25).

Die Grundlage der Verhaltensbeobachtung bilden je nach Rahmenbedingung qualitativ sowie quantitativ variierende Sachverhalte. Faßnacht bildet diese Tatsache im Datenschema der kaskadierten Prädikation ab.

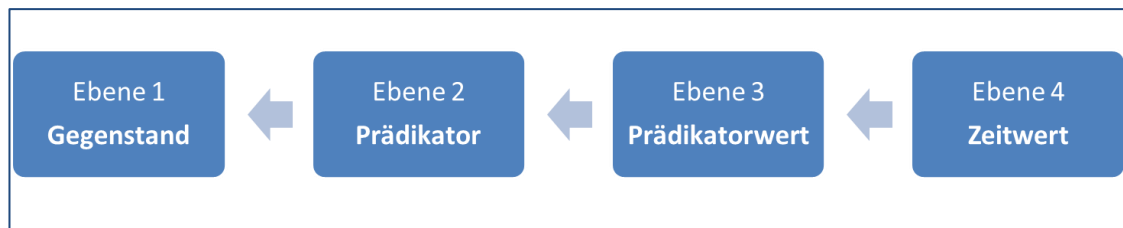


Abbildung 5. Datenschema als kaskadierte Prädikation (Faßnacht, 1995, S. 109)

Die 4 Ebenen Gegenstand, Prädikator, Prädikatorwert und Zeitwert können in der Genauigkeit ihrer Auflösung variieren. Wobei der Grad der Auflösung jedoch nicht die Objektivität der Verhaltensbeobachtung bedingt. Auf der Ebene der Gegenstandseinheiten, deren Verhalten beobachtet wird, unterscheidet Faßnacht (1995):

- *Einzelne Individuen,*
- *Dyaden,*
- *interagierende Gruppen,*
- *Menschenansammlungen*
- *Aggregat mehrere Individuen*

In der benachbarten Prädikatorebene werden bestimmte Eigenschaften gewählt, auf welche die jeweilige Gegenstandseinheit untersucht werden soll. Diese Eigenschaften müssen kontextunabhängig definiert werden und im Anschluss auf Ebene der Prädikatorwerte quantitativ und qualitativ bestimmt werden. Bei diesem Vorgehen werden Einheiten gewählt, welche eine empirische Wertmenge repräsentieren. Die letzte Ebene des kaskadierten Prädikatoren Schemas bildet der Zeitwert, welche ebenfalls in qualitative (Bsp. morgens oder abends) und quantitative (konkrete Zeitangabe) Zeitabschnitte aufgelöst werden kann (Faßnacht, 1995, S. 111-113).

Betrachtet man nun im Speziellen die Verhaltensquantifizierung, stößt man in der Literatur auf unterschiedliche Vorgehensweisen. Dabei zählen die Betrachtung von Latenzzeit, Häufigkeit, Dauer und Intensität zu den grundlegenden Messinstrumenten (Bateson & Martin, 1986, S.42-46). Je nach Verfahren werden diese Instrumente zur Verhaltensmessung unterschiedlich umgesetzt. Man unterscheidet:

- *Event-sampling-Verfahren*,
- *Time-sampling-Verfahren* und
- *Rating-Verfahren*

Unter *Event-sampling* versteht man die kontinuierliche Erfassung einer oder mehrerer Verhaltensweise, welche für einen längeren Zeitraum betrachtet und dabei möglichst genau erfasst werden sollen. Welches Verhalten im Fokus der Beobachtung steht, muss bereits vor der Untersuchung genau definiert und festgehalten werden. Aus einem Event-sampling resultieren quantitative Maße, wie Anfangs- und Endzeiten, Frequenzen, Verteilungen sowie absolute und relative Häufigkeiten.

Das *Time-sampling-Verfahren* ist ebenfalls eine gängige Methode zur Verhaltensquantifizierung. Hierbei werden die Beobachtungsgegenstände für eine festgelegte Zeit beobachtet. Dieses Vorgehen wird mehrere Male wiederholt, wobei die Verhaltensmessung in Form von Häufigkeits- und Zeitangaben realisiert wird.

Eine besonders leicht durchführbare Quantifizierungsmethode ist das *Rating-Verfahren*. Der Beobachter schätzt hierbei den Ausprägungsgrad der zu erfassenden Prädikatoren ein und überträgt diesen auf eine vorgegebene Skala. Es wird also dem, im Laufe der Beobachtung gezeigten Verhalten, je nach Ausprägung, ein bestimmter Skalenwert zugeordnet (Bateson & Martin, 1986, 40-46). So können in sehr kurzer Zeit auch komplexere Verhaltensweisen erfasst und in ihrer Ausprägung bewertet werden. Bei der Konstruktion von Ratingskalen müssen die drei folgenden Bedingungen zwangsläufig erfüllt werden. Zum einen sind die gewählten Kategorien exakt zu definieren. Dieses Kriterium wird auch Genauigkeitskriterium genannt. Des Weiteren ist auf die Einhaltung des Exklusivitätskriteriums zu achten. Die Kategorien dürfen sich demnach nicht überschneiden und müssen einander ausschließen. Das Exhaustivitätskriterium ist erfüllt, wenn die Kategorien die entsprechenden Merkmale erschöpfend beschreiben. Für Ausnahmen und nicht einordbares Verhalten kann eine Hilfskategorie „Sonstiges“ erstellt werden.

2.2.1 Güte einer ratingbasierten Verhaltensbeobachtung

Die Güte einer ratingbasierten Verhaltensbeobachtung hängt von zwei zentralen Faktoren ab. Zum einen müssen die potentiellen Beobachter genau geschult werden,

zum anderen sollten die eingesetzten Hilfsmittel den Güteansprüchen genügen (Hartmann, 1970, S. 176).

Um die Güte der Operationalisierung und des resultierenden Beobachtertrainings zu prüfen, müssen die folgenden Gütekriterien genauer untersucht werden.

Für eine gültige Datenerhebung muss unter anderem das Gütekriterium Objektivität gegeben sein. Eine Verhaltensbeobachtung ist objektiv, wenn sie die erhobenen Merkmale unabhängig von Testleiter, Auswerter und Interpretation misst (Moosbrugger & Kelava, 2008, S.9f). Ein Maß für die Objektivität ist die Interrater-Reliabilität, also das Ausmaß der Übereinstimmung verschiedener Beobachter. Der resultierende Koeffizient kann die Bewertungsgrundlage für die Qualität eines Beobachtertrainings bilden (Bortz & Döring, 1995, S.275ff). Um den Mehrwert eines Beobachtertrainings aufzuzeigen, wird die Bewertungsübereinstimmung vor und nach dem Training verglichen. Dabei gilt es zu untersuchen, ob sich die Interrater-Reliabilität durch die Weiterbildung erhöht.

Das Gütekriterium Reliabilität ist ein Maß für die Messgenauigkeit eines Verfahrens und gibt an, wie exakt das Verfahren misst. Mit Hilfe der internen Konsistenz können Rückschlüsse auf die Homogenität der Items, also auf die Kongruenz von Konstrukt und Indikatoren, gezogen werden. Zur Bestimmung der Retest-Reliabilität wird der gleiche Test zu 2 verschiedenen Testzeitpunkten durchgeführt und der Zusammenhang der Ergebnisse untersucht. Verfügt das Korrelationsmaß über eine bestimmte Mindestausprägung, misst das Verfahren annehmbar genau und darf eingesetzt werden. Es gilt zu prüfen, ob die gewählten Items homogen sind. Dies geschieht über die Betrachtung der Inneren Konsistenz, also Betrachtung der einzelnen Subskalen (Moosbrugger & Kelava, 2008, S.13).

Die Validität ist ebenfalls ein Maß für die Güte der Operationalisierung. Sie gibt an, ob die resultierenden Daten die Konstrukte, die das Verfahren erheben soll, auch tatsächlich repräsentiert. Die prognostische Validität wird sich in zukünftigen Drop-out-Raten zeigen.

2.2.2 Videovermittelten Verhaltensbeobachtung

Die Bearbeitung des DCT wird, nach Einverständniserklärung der Bewerber, mit einer Videokamera aufgezeichnet. Da die videobasierte Analyse des Bewerber-

verhaltens eine zentrale Rolle dieser Diplomarbeit spielt, soll nun auf methodische Vor- und Nachteile dieses Vorgehens genauer eingegangen werden.

Technisch vermittelte Beobachtungen sind in der heutigen Psychologie eine beliebte und somit weitverbreitete Erhebungsmethode, da sie eine exakte Analyse der vorliegenden Daten gewährleistet. Die Aufnahmen können für einen längeren Zeitraum gespeichert und somit jederzeit wiederholt, gestoppt oder verlangsamt werden (Greve & Wentura, 1997, S.27). All diese technischen Möglichkeiten haben zur Folge, dass selbst sehr schnelle und komplexe Vorgänge fassbar werden. Desweiteren können die Aufzeichnungen gespeichert und somit jederzeit repliziert werden. Außerdem ist die Ergänzung einer zusätzlichen Messmethoden, zum Beispiel der nachträglich Einsatz einer Stoppuhr, weitestgehend problemlos. Die Aufmerksamkeit des Beobachters kann selbst über einen längeren Zeitraum relativ konstant gehalten werden, da Pausen während der Analyse des Videomaterials kein Problem darstellen. Somit löst die Videoaufzeichnung zahlreiche Kritikpunkte der strukturierten Verhaltensbeobachtung (Faßnacht, 1993, S. 131-135).

Obwohl die videobasierte Erhebung viele Vorteile aufweist, stehen ihr auch einige Nachteile gegenüber. Im Gegensatz zur unvermittelten Beobachtung wird durch den Einsatz solcher Instrumentarien die systematische Ausblendung weniger relevanter Informationen erschwert. Aufgrund der mangelnden Dynamik kann ein zu enger oder starrer Blickwinkel einer statischen Videokamera wichtige Informationen verbergen (Greve & Wentura, 1997, S.27). Ein weiterer bekannter Nachteil der nicht verdeckten, speziell der videogestützten offenen, Verhaltensbeobachtung ist der sogenannte Hawthorne-Effekt. Dieser Effekt besagt, dass der Beobachter das Verhalten der Personen mit seiner bloßen Anwesenheit beeinflusst. Eine Lösungsmöglichkeit dieser Verzerrung stellt eine verdeckte Beobachtung dar. Dieses Vorgehen ist jedoch aus ethischen sowie rechtlichen Gründen im Kontext einer Bewerberauswahl ausgeschlossen. Das manifeste Verhalten während einer Beobachtung darf somit nur als Indiz für die latenten psychologischen Eigenschaften einer Person interpretiert werden (Bateson & Martin, 1986, S.70).

2.2.3 Verhaltensbeobachtung im Assessment Center

Gängige Persönlichkeitsverfahren sind sehr anfällig für Verfälschung. Bewerber befinden sich in einer Auswahl-situation und neigen zu sozial erwünschtem Antwortverhalten. Auch der Einsatz von Lügenskalen konnte diese Problematik bis zum jetzigen Zeitpunkt nicht lösen (Seiwald, 2003, 271 f). Verhaltensbeobachtungen, auch im Kontext eines Assessment Centers, verfügen, wie bereits erwähnt, über hohe Akzeptanz und prognostische Validität, da weniger im Sinne der sozialerwünschten Verhaltenstendenzen manipuliert werden können. Im Zeitraum von 1996 bis 2003 zeigte eine groß angelegte Langzeitstudie des DLR² signifikante Korrelationen zwischen dem Abschneiden in einem Assessment Center und dem späteren Ausbildungserfolg in der Flugführung (Pecena et al., 2013). Bei einem Assessment Center wird „... eine Teilnehmergruppe hinsichtlich einer Vielzahl von Dimensionen von mehreren trainierten Beobachtern gleichzeitig beurteilt“ (Hossiep, 2003, S.44). Die Bewerber bearbeiten komplexe Aufgaben in Kleingruppen und stehen dabei unter Beobachtung. Es zeigen sich Führungsverhalten, Interaktionsweisen und kommunikative Fähigkeit. Außerdem können spezifische Talente, Charakteristiken und Neigungen der Bewerber identifiziert werden. Anschließend wird das beobachtete Verhalten von speziell geschulten Beobachtern bewertet und entsprechend des Anforderungsprofils beurteilt (Hossiep, 2003, S.43ff). Die Verhaltensbeobachtung in Kontext eines Assessment Centers sollte zunächst mit Hilfe von speziellen Beobachtungsbögen protokolliert und erst nach Beendigung des Verfahrens ausgewertet werden. Hierzu werden die einzelnen Eindrücke in einem Gremium, der sogenannten Beobachterkonferenz, aufbereitet und diskutiert. Es erfolgt eine gemeinsame Beurteilung der Bewerberleistung, welche dem Kandidat auch persönlich rückgemeldet werden sollte.

2.3 Beobachtertraining

Die Bearbeitung des DCT soll von einer Kommission beobachtet werden. Diese setzt sich aus zwei bis drei Personen zusammen, mindestens einem Psychologen und einem praxiserfahrenen Fluglotsen. Die beobachteten Verhaltensweisen werden begleitend zur Testdurchführung notiert. Bewertung und Beurteilung der Bewerber werden erst im Anschluss in einer gemeinsamen Beobachterkonferenz diskutiert und

² Deutsches Zentrum für Luft- und Raumfahrt

festgelegt. Hossiep (2003) verweist in seiner Definition eines Assessmentcenter auf die Notwendigkeit von mehreren trainierten Beobachtern für eine gültige Bewerberbeurteilung. Diese spezielle Schulung soll mit Hilfe eines computergestützten Beobachtertrainings realisiert werden.

2.3.1 Beobachterfehler

Der DCT ist ein computergestütztes Messverfahren, bei welchem die Bewerber während der Bearbeitung von mehreren Personen beobachtet und anschließend bewertet werden. Der resultierende Messwert (X_t) ist, laut den Annahmen der klassischen Testtheorie, die Summe aus wahrem Wert (X_w) und einem Fehleranteil (ϵ).

$$X_t = X_w + \epsilon$$

Abbildung 6. Zusammensetzung Messwert nach Annahme der klassischen Testtheorie

Der Messfehler geht dabei auf das Messgerät zurück. Bei einer Verhaltensbeobachtung fungiert der Mensch als Messinstrument und ist somit auch Quelle möglicher Messfehler. Diese Annahme kann wie folgt verdeutlicht werden. Wird eine Person von zwei Beobachtern betrachtet, kann es durchaus vorkommen, dass diese zu unterschiedlichen Einschätzungen gelangen. Das *Verhaltens-Eindrucks-Aussage-Modell* von Brandstätter aus dem Jahr 1969 ist ein sozialpsychologischer Ansatz zur Beschreibung von sozialer Kognition, welche auch als soziale Urteilsbildung bezeichnet wird. Schuler (1996) bringt dieses Modell mit der Verhaltensbeobachtung im Kontext der psychologischen Personalauswahl in Verbindung und analysiert dabei bedeutende Elemente und Teilprozesse der Verhaltensbeobachtung. Differenzen der Beobachteraussagen werden durch drei Ebenen von Fehlerquellen erklärt. Demnach könnten verschiedene Verhaltensweisen beobachtet worden sein, dass gleiche Verhalten zu unterschiedlichen Eindrücken oder gleiche Eindrücke zu unterschiedlichen Beobachteraussagen geführt haben.

Ebene 1 – Das Verhalten

Je nach Situation kann das Verhalten von der Person selbst, deren Umgebung und durch die Wechselwirkungen dieser beiden Determinanten beeinflusst werden. Da

die Zielperson nur in bestimmten Situationen stichprobenartig beobachtet werden kann, darf keinesfalls von einer Repräsentativität der gezeigten Verhaltensweisen ausgegangen werden. Verschiedene Beobachtungszeitpunkte können somit unterschiedliche Verhaltensbeobachtungen zur Folge haben.

Ebene 2 –Der Eindruck

Verhalten ist in all seinen Facetten ein sehr komplexes Konstrukt und steht einer begrenzten Informationsverarbeitungskapazität auf Seiten des Beobachters gegenüber. Es können also immer nur bestimmte Verhaltensausschnitte registriert und verarbeitet werden. Die Informationsaufnahme der Beobachter ist eine selektive. Erkenntnisse der Wahrnehmungspsychologie gehen davon aus, dass interpretative Elemente nahezu immer Bestandteil der Wahrnehmung sind. Aus dieser Annahme kann abgeleitet werden, dass jede Beobachtung schon einen gewissen Grad an unbewusster Akzentuierung enthält (Hasemann, 1964, S.813). Die Auswahl der Beobachtungssegmente ist also stark subjektiv geprägt und wird unter anderem von Interesse, Fähigkeit, Erwartung, Einstellung, Bedürfnissen sowie der persönlichen Erfahrung bestimmt. Man unterscheidet hierbei kognitive, motivationale und emotionale Urteilsbedingungen.

Ebene 3 – Die Aussage

Eine weitere mögliche Ursache von Urteilsvarianzen ist die Überführung von Eindrücken in Beurteilungsaussagen. Die Transformation der Reize wird von der sprachlichen Kompetenz des Beobachters determiniert. Dies betrifft sowohl Syntax und Ausdrucksweise beim Niederschreiben als auch die Semantik beim Einordnen in Verhaltenskategorien. Auch bestimmte Strategien des Beobachters, wie die bewusste Datenfälschung, können dieser Ebene zugeordnet werden.

Da es sich bei der Bearbeitung des DCT um eine hoch standardisierte Situation handelt, welche von beiden Beobachtern gleichzeitig erfasst wird, rückt die Ebene eins in diesem Zusammenhang in den Hintergrund. In standardisierten Situationen sind Aufgabenverlauf und externe Einflüsse speziell für bestimmte Beobachtungszwecke konstruiert. Sie grenzen somit die Anzahl möglicher Verhaltensweisen ein. Die Aussage, dritte Ebene des *Verhaltens-Eindrucks-Aussage-Modells*, verliert durch eine gründliche Operationalisierung der Verhaltensdimensionen und die Entwicklung einer geeigneten Protokollieremethode an

Bedeutung (Hasemann, 1964, S. 811). Eine der Hauptquellen der Urteilsvarianz bildet jedoch die Ebene Eindruck. Die darin beschriebenen kognitiven, motivationalen und emotionalen Urteilsbedingungen bilden die Grundlage für typische Beobachtungs- oder Beurteilungsfehler.

Bei einer Verhaltensbeobachtung basieren die Messergebnisse auf die menschliche Wahrnehmungsfähigkeit. Dieses Vorgehen ist mit einigen Problemen verbunden. Je nach eigener Erfahrung und Beobachtungsfähigkeit können die Ergebnisse der Datenerhebung beeinflusst werden (Becker-Beck, 1997, S.114). Kommen unterschiedliche Beobachter bei gleicher Datenlage zu unterschiedlichen Resultaten, ist die Messung ungenau und entspricht nicht dem Gütekriterium Reliabilität. Eine Möglichkeit, wie es zu solchen unterschiedlichen Betrachtungsweisen kommen kann, sind die in der Literatur oft zitierten Beobachterfehler, welche vor allem auf der von Schuler (1996) angeführte Ebene des Eindrucks anzusiedeln sind. In Bezug auf die Beobachtung und Beurteilung im Kontext des DCT liegt der Fokus vor allem auf folgenden Fehlerquellen: Halo Effekt, Implizite Persönlichkeitstheorie, Primacy-/Recency-Fehler, Bezugsfehler, Similiar-to-me Effekt, Grundeinstellung des Beobachters und Dominanz einzelner Beobachter.

2.3.1.1 Halo Effekt

Der Halo-Effekt, welcher öfter auch als Hofeffekt bezeichnet wird, kommt dadurch zustande, dass die Wahrnehmung der Merkmale einer Person durch ein herausragendes Merkmal beeinflusst wird. Die Gesamtbeurteilung wird dadurch verzerrt, dass diese besonders positiven oder auch negativen Eigenschaften andere weniger ausgeprägte Merkmale überstrahlen (Amelang & Zielinski, 2002, S. 365).

2.3.1.2 Implizite Persönlichkeitstheorie

Dieser Beobachterfehler kann auch als logischer Fehler bezeichnet werden. Hier beeinflusst nicht eine herausragende Eigenschaft die anderen Merkmale einer Person, sondern die implizite Annahme über Zusammenhänge bestimmter Eigenschaften. Beobachter haben häufig vorgefasste Meinungen darüber, welche Persönlichkeitsmerkmale zusammen auftreten und welche einander ausschließen (Amelang & Zielinski, 2002, S. 365).

2.3.1.3 Primacy-/Recency-Fehler

Bei diesem Beobachterfehler hat der Zeitpunkt der wahrgenommenen Inhalte eine große Bedeutung. Sowohl der erste (primacy) als auch der letzte Eindruck (recency) kann einen Einfluss auf die Gesamtbeurteilung haben und diese somit verzerren. Primacy- und Recency-Effekte schließen sich jedoch nicht aus und können durchaus auch zusammen auftreten (Faßnacht, 1993, S. 221).

2.3.1.4 Bezugsfehler

Dieser Fehler bezieht sich auf den Bezugsrahmen einer Beurteilung. Leistungen werden demnach nicht absolut, sondern in Bezug zu den anderen Personen bewertet. Es könnte also zum Beispiel ein durchschnittlich geeigneter Bewerber in einer Gruppensituation mit überdurchschnittlichen Bewerbern als eher unterdurchschnittlich beurteilt werden (Eschen-Léguedé, 2012, S. 11).

2.3.1.5 Similiar-to –me-Effekt

Einige Beurteiler neigen dazu, persönliche Interessen, Fähigkeiten, Eigenschaften oder Wünsche dem Bewerber zuzuordnen. So werden diejenigen Personen positiver beurteilt, von denen der Beobachter annimmt, dass sie ihm ähnlich sind. Die nicht zu vernachlässigende Gefahr dieser Projektionsvorgänge liegt darin begründet, dass sie meist unbewusst und damit unkontrolliert ablaufen (Faßnacht, 1993, 223).

2.3.1.6 Grundeinstellung des Beobachters

Je nach Grundeinstellung des Beobachters unterscheidet man zwischen der Tendenz zu zentralen, milden, strengen sowie Extremurteilen. Ein Mildefehler zeichnet sich dadurch aus, dass ein Beobachter insgesamt bessere Urteile abgibt und somit im Vergleich zu anderen Beobachtern höhere Mittelwerte erzielt. Ein Strengfehler beschreibt im Gegensatz dazu die Tendenz zu unterdurchschnittlichen Urteilen. Liegt eine vergleichsweise geringe Varianz der Beurteilungen vor, spricht man von zentralen Tendenzen eines Beobachters. Eine erhöhte Varianz der Urteile spricht hingegen für eine Tendenz zu Extremurteilen (Amelang & Zielinski, 2002, S. 365).

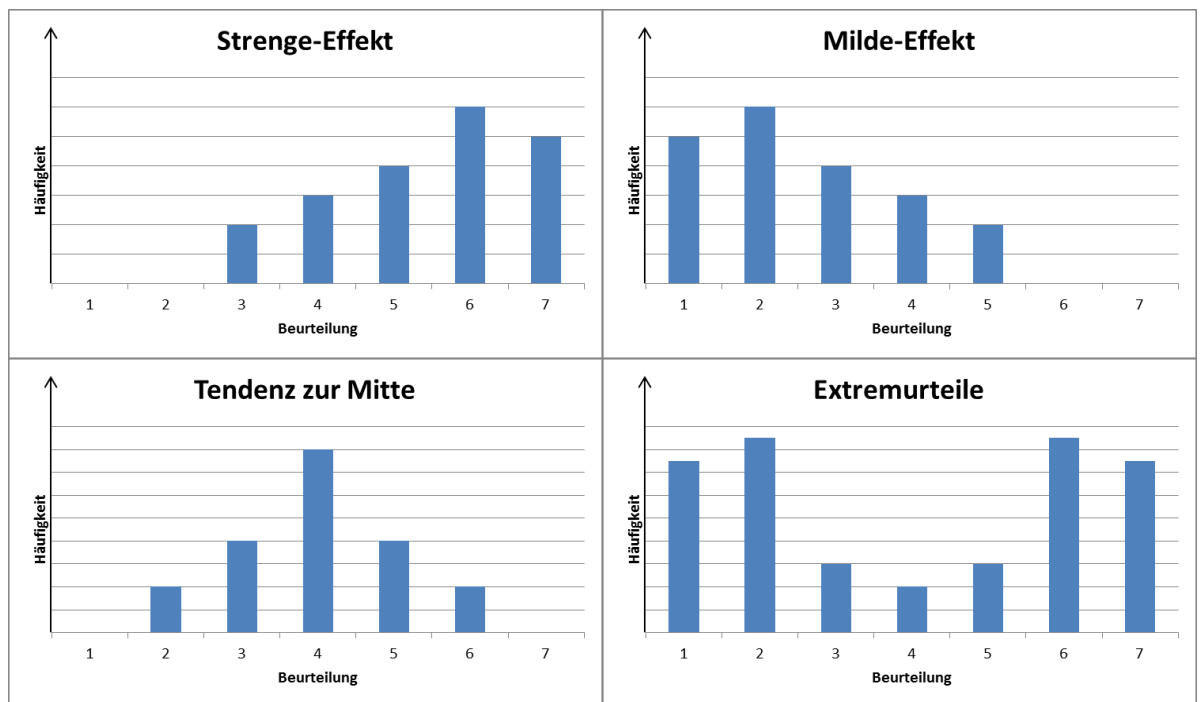


Abbildung 7. Illustration Grundeinstellung des Beobachters (nach Amelang & Schmidt-Atzert, 2012. S.321)

2.3.1.7 Dominanz einzelner Beobachter

Kommt es im Anschluss an eine Verhaltensbeobachtung zu einer Beobachterkonferenz, kann es zur Dominanz einiger weniger Beobachter kommen. Bei der Bewertungs- und Entscheidungsdiskussion treten diese besonders stark in den Vordergrund. Diese Beobachter bringen ihre Ansichten sehr oft und ausführlich ein, während andere ihre Eindrücke zurückhalten und wenn, nur sehr zögerlich widersprechen (Eschen-Léguedé, 2012, S. 12).

2.3.2 Inhalte und Ziele des Beobachtertrainings

Da es sich bei einer Bewerberauswahl um ein folgenschweres Verfahren mit enormer Tragweite handelt, sollte eine Diagnose keinesfalls auf der natürlichen Beobachtungsgabe des Menschen basieren. Wie alle anderen Messinstrumente wird auch der Mensch nie vollkommen exakt urteilen können. Mit Hilfe der Durchführung eines speziellen Beobachtertrainings können die oben angeführten Beobachtungs- und Beurteilungsfehler zumindest gemildert werden. Nach Hasemann (1964) dient eine Schulung von potentiellen Beobachtern der Gewöhnung an die Auffassungsart

durch Übung. Hasemann postuliert folgende methodischen Grundbausteine eines Beobachtertrainings:

1. Schulung der Wahrnehmungsschärfe

Hierbei soll die Beobachtung und Wiedergabe der gewonnen Eindrücke trainiert und durch fortlaufendes Feedback modifiziert werden. Für dieses Vorgehen werden Filmaufnahmen empfohlen.

2. Übung der Beobachtung nach Teilaspekten

Die Hauptinhalte bilden hier die Vermittlung der beobachtbaren Dimensionen und deren Zerlegung in Einzelteile. Dabei sollt jedoch der Zusammenhang des Ganzen nie aus den Augen verloren werden.

3. Übung der rein sprachlichen Sicherheit des Beschreibens

Diese Fertigkeit kann durch freie mündliche Erzählung ganzer Beobachtungseinheiten oder bestimmter Teilaspekte trainiert werden.

Kleinmann (2003) überträgt diese Grundlagen auf ein Beobachtungstraining im Kontext von Assessment Centern und schlussfolgert folgende Hauptkategorien, welche sowohl in Kombination als auch isoliert Anwendung finden können:

1. Beobachterfehlertraining

Anhand von Stimulusmaterial wird auf gängige Beobachterfehler aufmerksam gemacht, mit der Zielsetzung diese durch die gezielte Bewusstmachung zu vermeiden.

2. Dimensionsverwendungstraining

In dieser Kategorie des Beobachtertrainings wird geübt, gezeigtes Verhalten (schriftlich oder Videoaufnahme) bestimmten Verhaltensdimensionen zuzuordnen.

3. Bezugsrahmentraining

Hierbei wird das Verhalten möglichst realistisch dargeboten und von den Schülern bewertet. Durch die ständige Rückmeldung soll der Umgang mit den Beurteilungsskalen verinnerlicht werden.

4. Verhaltensbeobachtungstraining

Die Trennung von Beobachtung und Beurteilung steht im Fokus dieser Kategorie. In den Übungen muss das gezeigte Verhalten erst einer Dimension zugeordnet und erst im Anschluss mit Hilfe der entsprechenden Skala bewertet werden.

Wie Meta-Analysen zeigen konnten, ging die Fehlerrate durch eine gezielte Schulung der Beobachter zurück. Halo-Effekte und Mildefehler konnten reduziert werden, wohingegen Beobachtungs- und Beurteilungsakkuratheit infolge des Trainings anstiegen (Huffcutt & Woehr, 1994).

Resümierend ist festzuhalten, um die Objektivität einer Verhaltensbeobachtung sicherzustellen, ist ein spezielles Beobachtertraining von dringender Notwendigkeit. Das Training sollte die für die Verhaltensbeobachtung relevanten Inhalte effizient und anschaulich vermitteln. Die zu erarbeitende Hauptinhalte bilden in Anlehnung an bestehende Trainingssysteme:

- *Verhaltensbeobachtungstraining,*
- *Dimensionsverwendungstraining,*
- *Bezugsrahmentraining und*
- *Beobachterfehlertraining* (Eschen-Léguedé, 2006).

Ein Verhaltensbeobachtungstraining vermittelt die grundlegende Vorgehensweise einer wissenschaftlichen Beobachtung. Hier kann zusätzlich mit Hilfe von Videobeispielen die Beobachtungssituation geübt bzw. trainiert werden. Bestimmte Verhaltensweisen anhand von Beschreibungen den entsprechenden Konstrukten zuzuordnen, wird in einem Dimensionstraining geschult. Im Anschluss dazu wird das zu beobachtende Verhalten in einem Bezugsrahmentraining an realistischen Beispielen dargeboten und so die entsprechende Quantifizierung, also die richtige Einteilung im Bewertungssystem erläutert. Im Beobachterfehlertraining werden typische Fehler der Verhaltensbeobachtung mit dem Ziel einer Reduzierung selbiger demonstriert. Es geht also bei diesem Training um die Vermittlung von Ablauf, Dimensionen, Indikatoren und häufige Beobachterfehler. Dabei können die inhaltlichen Schwerpunkte mittels interaktiver Lehrmaterialien aufbereitet und

oftmals auftretende Verhaltensmuster via Videomaterial aus den Testdurchläufen demonstriert werden.

3. FRAGESTELLUNG

Aus gegebener Notwendigkeit müssen die interessierenden Dimensionen des DCT operationalisiert werden. Um die Qualität der erarbeiteten Operationalisierung fundiert begründen zu können, muss die Güte der resultierenden Daten sowie eine Qualitätssteigerung durch das Beobachtertraining belegt werden. Dies führt zu den folgenden Hauptfragestellungen und den entsprechenden Hypothesen dieser Diplomarbeit.

Forschungsfrage 1 – Verwendbarkeit der Aufgabe

Liefert die videobasierte Operationalisierung im Kontext des DCT psychometrisch verwendbare Daten?

- **Hypothese 2.1**

Das DCT-Verfahren erfüllt das Gütekriterium Objektivität. Die Koeffizienten der Interrater-Reliabilität liefern zufriedenstellende Werte. $r \geq 0,4$

Forschungsfrage 2 – Evaluation des Beobachtertrainings

Führt das Beobachtertraining auf Grundlage der videobasierten Operationalisierung zu einer Verbesserung der Verwendbarkeit der psychometrische der Daten?

- **Hypothese 2.1**

Das Training führt zu einer Erhöhung der Beobachterübereinstimmung. Die Koeffizienten der Interrater-Reliabilität erhöhen sich. $r_{t1} < r_{t2}$

- **Hypothese 2.2**

Das Training führt zu einer Reduktion von Beobachterfehlern. Zentrale Tendenzen, Milde- bzw. Strengeeffekte werden geringer. $\varepsilon_{t1} < \varepsilon_{t2}$

4. METHODIK

Das vierte Kapitel dokumentiert alle Bearbeitungsschritte, um die oben angeführten Fragestellungen fundiert beantworten zu können. Die Dimensionen Belastbarkeit, Arbeitsstil und Interaktion wurden bereits im Theorieteil genau definiert. Nun gilt es, die Konstrukte in den Kontext des DCT zu übertragen. Im Abschnitt Operationalisierung wird die Generierung der Indikatoren, deren Klassifikation und Quantifizierung genau beschrieben. Auch das Vorgehen zur Erarbeitung eines Beobachtertrainings und die Methodik der Datenerhebung werden im nun folgenden Kapitel detailliert erläutert.

4.1 Operationalisierung

In diesem Abschnitt wird auf den konkreten Ablauf der Videoanalyse explizit eingegangen. Nach der theoretisch fundierten Bedeutungsanalyse liegt nun eine analytische Definition der Konstrukte vor, welche im nächsten Schritt zusätzlich operational definiert und spezifiziert werden sollen. Für die Angabe der Operationen zur Erfassung der psychologischen Konstrukte benötigt man passende Verfahren und Messmethoden, welche für jeden Arbeitsschritt nachvollziehbar dokumentiert werden. Die Operationalisierung setzt sich dabei wie folgt zusammen. Aus den typischen Verhaltensweisen, welche im Laufe der DCT-Bearbeitung beobachtbar sind, werden Indikatoren abgeleitet (Börtz & Döring, 1995). Diese Verhaltensanker werden gelistet, strukturiert und zu Ereignisklassen zusammengefasst, um diese im Anschluss mit Hilfe transparenter Zuweisungsregeln auf einer siebenstufigen Skala zu quantifizieren.

4.1.1 Generierung der Indikatoren

Für die Konstruktion eines psychologisch diagnostischen Testverfahrens bilden fundierte Kenntnisse des Messgegenstandes eine unerlässliche Voraussetzung (Smidt-Atzert, Kersting, Preckel, Westhoff & Ziegler, 2012). Im vorherigen Theorieteil wurden die, aus Anforderungsanalysen und theoretischen Überlegungen resultierenden Konstrukte mit Hilfe aktueller Fachliteratur genau definiert, spezifiziert und von themenverwandten Inhalten abgegrenzt. Durch diese umfassende

Literaturrecherche wurde ein bedeutender Grundstein gelegt, der nun in den Kontext des DCT übertragen werden muss. Indikatoren bestehen aus typischen Verhaltensweisen, welche von den Akteuren während der Bearbeitung gehäuft gezeigt werden. Um diese Verhaltensanker zu generieren, reichen theoretische Überlegungen nicht aus. Das Verhalten der Probanden muss in einer größeren Zahl von Testläufen genau erforscht werden. Die hierfür benötigten Testpersonen entspringen der Gruppe der Bewerber für die FFD. Durch eine obligatorische Durchführung des DCT, im Anschluss an das gängige Auswahlverfahren am Zentrum für Luft- und Raumfahrtmedizin der Bundeswehr, konnte eine größere Anzahl an DCT-Durchläufen im Detail betrachtet werden.

Die Anfertigung von Videoaufnahmen der DCT-Bearbeitungen ermöglichte eine detaillierte und replizierbare Analyse der gezeigten Verhaltensweisen. Zur Generierung der Verhaltensanker wurde das Videomaterial von insgesamt zehn verschiedenen Testdurchläufen ausgewertet. Durch die gemeinsame Bearbeitung des Szenarios entspricht dies der Sichtung von insgesamt 20 Testpersonen. Die Bewerber der FFD stimmten vor der Bearbeitung durch eine Einverständniserklärung der Video- und Tonaufzeichnung sowie der geplanten Nutzung für interne Schulungszwecke schriftlich zu. Vor jedem Bearbeitungsdurchgang informierte ein geschulter Testleiter alle Teilnehmer über den Nutzen und die anonymisierte Verwendung des Aufnahmematerials. Außerdem wurde instruiert, dass das Abschneiden keinerlei Einfluss auf die laufende Bewerbung hat und die erbrachte Testleistung im anschließenden Psychologeninterview nur rückgemeldet wird, falls dies der ausdrückliche Wunsch des Bewerbers sei. Die Durchläufe wurden aus Sicht der Beobachterkommission, also von der Rückseite, gefilmt. Somit sind die Personen auf dem Videomaterial nicht frontal zu sehen. Durch diese Perspektive und den Einsatz von Probandencodes wurde die Anonymität der Teilnehmer für alle Auswertungsschritte sichergestellt. Dieser Blickwinkel barg den Nachteil, dass so nahezu alle Informationen über den Gesichtsausdruck verloren gingen. Neben der Anonymität bildete ebenfalls der Ausblick auf den Bearbeitungsbildschirm, unerlässlich für eine nachvollziehbare Verhaltensbeobachtung, schwerwiegende Argumente für die bewusste Wahl der Rückenperspektive.

Im Zeitraum von 29.10.2013 bis 07.11.2013 wurde der gesamte DCT, paarweise von drei Frauen und siebzehn Männern bearbeitet. Eine Videoaufzeichnung erfolgte

dabei ausschließlich für die sechs Teamaufgaben, welche von den Partnern gemeinsam bearbeitet wurden. Dieses Segment wurde aus dem Grund gewählt, da die Priorität des DCT auf Verhaltensweisen im Kontext von Interaktionssituationen liegt. Die Bearbeitungszeit des Teamteils variiert durch die selbstgewählten Start- und Endpunkte, befand sich im Durchschnitt jedoch in etwa bei 25 Minuten. Die aufsummiert 250-minütigen Videomaterialien wurden im Anschluss an die Erhebungsphase tabellarisch transkribiert. Der Handlungsverlauf der beiden Akteure wurde einander in zwei parallel verlaufenden Spalten gegenübergestellt. Eine weitere dritte Spalte hatte den Verwendungszweck sonstiger Vermerke und zusätzlicher Notizen. Dabei wurde darauf geachtet, jegliche Art der Deutung oder Interpretation zu unterlassen. Für jede der sechs Runden wurden alle verbalen sowie nonverbalen Verhaltensweisen dokumentiert. Da aus perspektivischen Gründen die Mimik der Personen nur in wenigen Einzelfällen beobachtet werden konnte, wurden vordergründig mündliche Aussagen, Bewegungen und Computereingaben der Bewerber transkribiert. Aufgrund der großen Datenmenge und der sehr standardisierten Bearbeitungssituation wurden alle Informationen paraphrasiert. Außerdem wurden alle Aussagen in normales Schriftdeutsch transformiert. Auf Basis einer gründlichen Analyse des erstellten Materials, erfolgte die Abstraktion der gezeigten Verhaltensweisen, also eine Reduktion auf die wesentliche Bedeutung. Anschließend wurden die abstrahierten Verhaltensweisen gelistet.

4.1.2 Klassifikation der Indikatoren

Nach Selektion und Abstraktion der gezeigten Verhaltensweisen folgte die Klassifikation der beobachteten Inhalte. Die Vielzahl der möglichen Handlungen mussten strukturiert und theoretischen Ereignisklassen zugewiesen werden. Diese Klassen fassen Ereignisse mit ähnlicher Bedeutung zusammen (Bortz & Döring, 1995, S.266). Zur Strukturierung und Kategorisierung des umfangreichen Transkripts wurde die QDA-Software MAXQDA 11 verwendet. Die Kategorienbildung verlief im ersten Bearbeitungsschritt deduktiv. Anhand der entsprechenden Literatur wurden die Hauptkategorien Belastbarkeit, Arbeitshaltung, Interaktion und Kognitive Leistung gewählt. Insgesamt wurden den verschiedenen Kategorien 1224 Verhaltensanker zugewiesen. Der nächste Arbeitsschritt beinhaltete die induktive Erweiterung dieser vier Hauptkategorien, also die Konstruktion diverser

Subkategorien, aus dem Material selbst. Bei der Benennung dieser Subkategorien wurde themenspezifisch Bezug auf die Basisliteratur genommen. Es resultierte folgendes Codiersystem (Abbildung 8), welches einem hierarchischen Aufbau unterliegt

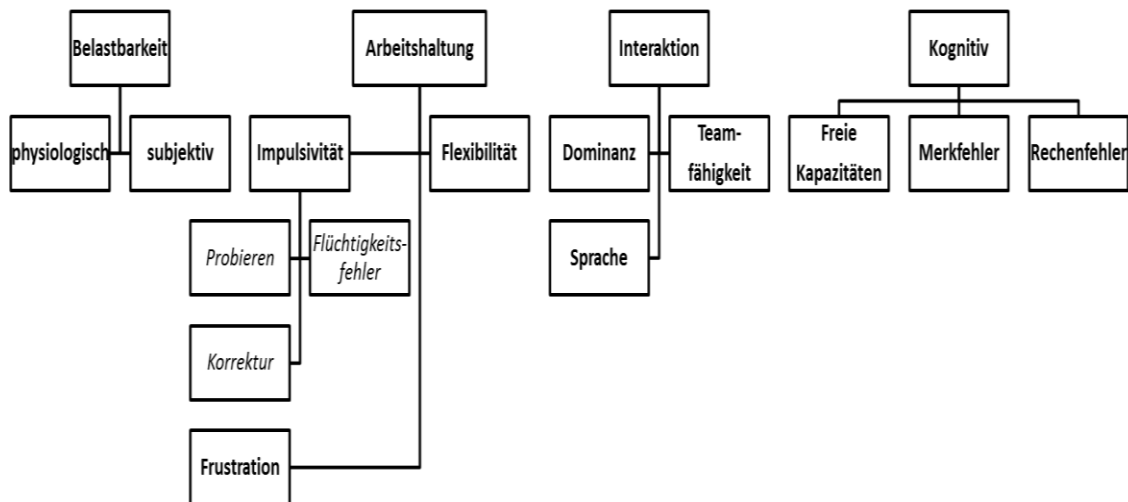


Abbildung 8. Codiersystem der qualitativen Inhaltsanalyse des Videomaterials

Für eine nachvollziehbare Bewertung des DCT fungiert Transparenz als Kardinalkriterium. Jedes Verhalten das in die Bewertung eingeht, muss eindeutig einer Kategorie zugewiesen werden können. In einer hoch standardisierten Situation, wie dem DCT-Szenario, sind Häufungen bestimmter Verhaltensweisen unabdingbar. Für die Kategorisierung ist jedoch nur das Vorhandensein, nicht die Häufigkeit einer Verhaltensweise von Bedeutung. Aus diesem Grund wurden alle mehrfach gezeigten Verhaltensweisen zu einem einzigen Anker zusammengefasst. Durch diese Selektion konnten die ursprünglichen 1224 Verhaltensweisen auf 263 Indikatoren reduziert werden. Listen mit den Verhaltensindikatoren und deren Zuordnung zu den entsprechenden Subkategorien bilden die Grundlage des Gütekriteriums der Nachvollziehbarkeit und befinden sich im Anhang.

4.1.2.1 Klassifikation Belastbarkeit

Als theoretische Basis der Kategorie Belastbarkeit fungiert die Einteilung von Wieland-Eckelmann (1992). Im Fokus stehen somit physiologische und subjektive Belastungsreaktionen während des gesamten Bearbeitungszeitraums. Typische physiologische Verhaltensanker sind nach Schwarzer (1981) motorische Unruhe, wie

zum Beispiel unruhige Beine, trommelnde Finger und körperliche Anspannung. Diese Gespanntheit kann durch Verhaltensweisen wie Zittern oder eine verkrampfte Körperhaltung beobachtet werden.

Die Subkategorie subjektive Belastungsanzeichen untergliedert sich laut Gros (1994) in Emotion, Konzentration, Koordination und Wahrnehmung. Der emotionale Belastungsaspekt wird in den meisten Fällen mündlich mit Äußerung von Unsicherheit, Selbstvorwürfen oder einem verlegenem Lächeln kommuniziert. Die Konzentration betreffende Beeinträchtigungen zeigen sich durch lautes Rechnen und den Gebrauch der Finger für einfache Berechnungen und zum Merken der Maximalbelastung. Bedient der Bewerber den Nummernblock mit beiden Händen oder drückt er die Tasten sehr fest, kann auf koordinative Belastungsanzeichen geschlossen werden. Der letzte Unterpunkt bezieht sich auf die Wahrnehmung. Häufig beobachtete Verhaltensanker dieser Subkategorie setzen sich aus einer sehr geringen Distanz zum Computerbildschirm und das Verfolgen des Bildschirmtextes mit dem Finger zusammen.

4.1.2.2 Klassifikation Arbeitsstil

Die Hauptkategorie Arbeitshaltungen konnte mit Hilfe der induktiven Vorgehensweise ebenfalls in weitere Subkategorien aufgegliedert werden. Die Unterpunkte Impulsivität, Regelkonformität und Frustration sollen zur Spezifizierung des Oberbegriffs beitragen. Impulsivität, in Anlehnung an Kagan (1965), beschreibt die Art und Weise, wie ein Akteur das DCT-Szenario bearbeitet. Eine lange Eingabedauer und das Kreisen mit dem Finger über dem Nummernblock sind Verhaltensanker, welche für eine zögerliche Bearbeitungsweise sprechen. Viele Korrekturen, egal ob mündlich oder mit Hilfe der Korrekturtaste, und Flüchtigkeitsfehler, wie das Drücken einer falschen Taste, sprechen für eine unüberlegte, spontane Form der Bearbeitung. Auch Verhaltensanker vorschneller Handlungen, also das Rechnen während einer Meldung und die Bearbeitungsstrategie des Probierens, wie die Beispielaussage: *"Ich versuche irgendwie zu verteilen, bis eine Fehlermeldung kommt."*, können der unüberlegten Bearbeitungsform zugeteilt werden. Der DCT ist so programmiert, dass es während der Bearbeitung zwangsläufig zu frustrierenden Situationen kommt. Die Durchgänge

fünf und sechs sind nicht lösbar, da hier in einer Runde zu wenige, und in der anderen Runde dem Verhältnis widersprechende Fahrzeugtypen vorhanden sind. Die Bearbeitungszeit ist mit zwei bis vier Minuten pro Runde absichtlich sehr knapp bemessen, gleichzeitig werden die geforderten Rechenoperationen und Merkleistungen zunehmend anspruchsvoller. Resultate hiervon sind häufige Fehlermeldungen aus Überlastungsgründen oder das Ablaufende der Bearbeitungszeit. Dies geschieht im Laufe der Bearbeitung mit zunehmender Wahrscheinlichkeit. Wie die Bewerber mit diesen Rückschlägen umgehen, beschreibt die Subkategorie Frustrationstoleranz. Theoretisch orientiert sich diese Kategorie an der Definition von Karner und Wagner aus dem Jahr 2007. Die Verhaltensanker reichen von „*bleibt ruhig und entspannt*“, bis Fluchen und Aufstöhnen bei Fehlermeldungen oder abgelaufener Bearbeitungszeit. Auch die Flexibilität in Anlehnung an Schmuck (1996) ist eine Subkategorie der Interaktion. Im Kontext des DCT-Szenarios ist Regelkonformität die passendere Formulierung. In der Instruktion werden die Bearbeitungsregeln des DCT verkündet. So heißt es, dass die 1. Meldung stets mit dem Wortlaut: „*Hier spricht City Ost/West.*“, mit der Nennung der eigenen Verantwortlichkeiten und aller Sondermeldungen, in denen das Wort „Problem“ oder „problematisch“ vorkommt, beginnen muss. Die Verhaltensanker, welche die Einhaltung dieser drei Regeln beschreiben, unterliegen der Subkategorie Regelkonformität. Aus der Aufgabenstellung des DCT ergibt sich die große Bedeutung der Rollenübernahme. Im Laufe der Bearbeitung ist man für bestimmte Straßen verantwortlich, übernimmt die Rolle des Koordinators oder dessen Zuarbeiters. Um sich diesen variablen Anforderungen anpassen zu können, bedarf es eines bestimmten Maßes an Flexibilität. Indikatoren, ob die Anpassung an eine Rolle gelingt oder nicht, wurden der Definition der jeweiligen Aufgabenbereiche, wie sie auch in der Instruktion des DCT zu finden sind, abgeleitet. So trägt ein Koordinator die Verantwortung und muss den Zuarbeiter anleiten. Im Gegensatz dazu muss dieser die Anweisungen ausführen und hat somit eine unterstützende Funktion. Würde ein Zuarbeiter zum Beispiel selbst über seine nächsten Schritte entscheiden, ohne dies mit dem Koordinator abzusprechen, widerspräche dies der Rollendefinition. Auch der Umgang mit neuen Situationen, wie Systemausfall oder die unlösbaren Aufgaben fünf und sechs, sind wichtige Verhaltensanker der Subkategorie Flexibilität.

4.1.2.3 Klassifikation Interaktion

Die dritte Hauptkategorie Interaktion setzt sich aus den Unterpunkten Dominanz, Kooperation und Kommunikation zusammen. Dominanz bezieht sich auf die soziale Kontrolle einer Person und befindet sich im Zusammenhang mit Becker-Becks dyadischen Interaktionsmuster, Kontrolle und Dominanz. Im Fokus steht somit, inwieweit ein Akteur seine eigenen Intentionen umsetzt und wie mit den Anweisungen des Partners umgegangen wird. Unterwürfige Verhaltensanker sind alle Formen des selbstlosen Punkteverzichts. Auch Handlungen der Absicherung, wie das Erfragen der Richtigkeit eigener Berechnungen, sind der Hauptkategorie Interaktion unterstellt. Im Gegensatz zur Subkategorie Unterwürfigkeit stehen bestimmende Verhaltensweisen. Beispiele hierfür sind Anweisungen im Imperativ und das Handeln ohne Absprache. In der Subkategorie Kooperation beziehungsweise Teamfähigkeit steht die Art der gemeinsamen Bearbeitung im Fokus. Grundlage dieser Kategorie ist das gleichnamige Interaktionsmuster nach Becker-Beck (1997). Verläuft eine Entscheidungsfindung im Team oder arbeitet jeder in Einzelarbeit für sich alleine? Aktive Hilfe, wie Alternativvorschläge, Äußerungen konstruktiver Kritik sowie Absprache über Strategie und Eingaben, sind zentrale Indikatoren der Teamarbeit. Ein weiteres Fragment der Interaktion bildet die Subkategorie Kommunikation, basierend auf der Einteilung von Fleischmann (2010). Ausdruck, Sprachverständnis und Informationsfilterung bilden hierbei die zentralen Punkte. Die Anker beziehen sich zum einen auf die mündliche Ausdrucksweise, *„Der Bewerber spricht laut und deutlich.“*, und zum anderen auf die Struktur der gesprochenen Inhalte, *„Die Anweisungen sind sehr knapp und präzise.“* oder *„Der Bewerber rechnet die Verhältnisse gleich um.“*. Das Augenmerk liegt auf der Effektivität der Kommunikation, also inwieweit werden Informationen verständlich aufbereitet, verkürzt und vom Partner richtig verstanden.

4.1.2.4 Klassifikation kognitive Fähigkeiten

Obwohl die Erhebung der kognitiven Fähigkeiten im DCT nur eine untergeordnete Rolle spielt, wäre eine völlige Vernachlässigung diesbezüglicher Informationen unökonomisch und mindert den diagnostischen Mehrwert des Verfahrens. Entsprechend der nebensächlichen Bedeutung wird erstmalig an dieser Stelle und nur kurz auf das Konstrukt eingegangen. Freie Kapazitäten der Bewerber zeigen sich,

wenn neben den obligatorischen Bearbeitungsschritten, auch ein Teil der Aufmerksamkeit auf Zeitmanagement und Punktwerte der Fahrzeuge gelegt wird. Nachdem die Maximalbelastungen vor jedem Durchgang nur 30 Sekunden dargeboten werden, können mit Hilfe der anschließenden Bearbeitung wichtige Rückschlüsse auf die Merkfähigkeit des Bewerbers gezogen werden. Durch den Umgang mit Einschränkungen und Verhältnissen in der Straßenbelegung lassen sich ähnliche Folgerungen auch in Bezug auf die Rechenfähigkeit ziehen.

4.1.3 Quantifizierung der Indikatoren

Die gezeigten Verhaltensweisen im Laufe der DCT-Bearbeitung wurden gesammelt, strukturiert und klassifiziert. Durch diese Vorgehensweise konnte ein hierarchisches Netz aus vier Hauptkategorien, neun Subkategorien und insgesamt 263 Verhaltensankern gebildet werden. Der anschließende Arbeitsschritt umfasste die Quantifizierung dieser Handlungsweisen. Den Indikatoren wurden in Bezug auf die jeweilige Dimension bestimmte Vorzeichen zugeschrieben. Eine Verhaltensweise kann somit Anker für eine positive beziehungsweise negative Ausprägung eines Konstrukts darstellen. Für eine eindeutige Quantifizierung müssen Häufigkeit und Intensität der Verhaltensanker nach genauen Regeln bewertet werden. In der Eignungsfeststellung des Dezernats Auswahl nutzt man für die Bewertung der bisherigen Eigenschaften eine siebenstufige Skala, die Bundeswehr-Notenskala. Hierbei stellt die Note 1 die beste, Note 7 die schlechteste Ausprägung eines Merkmals dar. Auch in Bezug auf den DCT soll an dieser Einteilung festgehalten werden, um somit eine einheitliche Auswertung der Ergebnisse gewährleisten zu können. Wie in den vorherigen Arbeitsschritten ist auch hier eine transparente Zuordnung unabdingbar. Die in den Videoaufnahmen gezeigten Verhaltensanker wurden zum Teil durch theoretische Überlegungen ergänzt. Dies war vor allem für die Besetzung der Extremausprägungen notwendig. Die theoretisch konstruierten Indikatoren sind in den bereits erwähnten Indikatorlisten im Anhang gesondert gekennzeichnet.

Diese Quantifizierung bildet die Bewertungsgrundlage des DCT. Entsprechend der Grundlagen einer wissenschaftlichen Verhaltensbeobachtung sollen Beobachtung und Beurteilung in zwei verschiedenen Schritten getrennt voneinander vorgenommen

werden. Gegenstand der Beobachtung sind dabei die beiden Bewerber, also ein dyadisches System. Die zu beobachtenden Prädikatoren setzen sich aus Belastbarkeit, Arbeitsstil, Interaktion und kognitive Leistung, inklusive der entsprechenden Subkategorien zusammen. Im Zeitintervall der sechs Spielrunden werden alle relevanten Verhaltensweisen erhoben und für jeden Bewerber getrennt notiert. Erst im Anschluss an die Verhaltensbeobachtung erfolgt die Bewertung, also die Interpretation der Daten. Ausgehend von den folgenden Zuweisungsregeln müssen die gezeigten Indikatoren den jeweiligen Subkategorien zugeordnet und mit der entsprechenden Note bewertet werden.

4.1.3.1 Zuweisungsregeln Belastbarkeit

Physiologische Belastungsreaktion

Bei der physiologischen Belastungsfähigkeit muss unterschieden werden, ob die Verhaltensanker, welche der Indikatorliste entnommen werden können, handlungsbegleitend oder handlungssteuernd sind. Sind die Indikatoren wahrnehmbar, haben jedoch offensichtlich keinen negativen Einfluss auf die gezeigte Performance, besitzen sie eine handlungsbegleitende Ausprägung. Sind aufgrund der wahrgenommenen Indikatoren starke Leistungseinbußen zu verzeichnen oder eine Aufgabenbearbeitung nicht mehr möglich, sind diese als handlungssteuernd einzuordnen. Denkbare Beispiele hierfür sind starkes Zittern, welches die Eingabe via Nummernpad unmöglich macht, starke Einschränkung der Sprachfähigkeit durch Stottern oder Ohnmacht. Neben Ausprägungen und Intensität der Verhaltensweisen ist auch die Anzahl der physiologischen Belastungsanzeichen eine wichtige Quantifizierungsregel. Dabei ist zu beachten, die Anzahl bezieht sich auf die Runden, in denen physiologische Belastungsanzeichen beobachtet werden können, nicht auf absolute Zahlen. Sobald in einer der sechs Bearbeitungsrounden handlungssteuernde physiologische Belastungsanzeichen gezeigt werden, kommt es zur Vernachlässigung der handlungsbegleitenden physiologischen Anzeichen.

Tabelle 3. Quantifizierung physiologische Belastungsreaktion

1	Keine physiologischen Belastungsanzeichen
2	In 1-2 Runden handlungsbegleitende physiologische Belastungsanzeichen (wahrnehmbar, Leistung wird nicht negativ beeinflusst)
3	In 3-5 Runden handlungsbegleitende physiologische Belastungsanzeichen
4	In jeder Runde handlungsbegleitende physiologische Belastungsanzeichen
5	In 1-2 Runden handlungssteuernde physiologische Belastungsanzeichen (so stark ausgeprägt, dass Leistung negativ beeinflusst)
6	In 3-5 Runden handlungssteuernde physiologische Belastungsanzeichen
7	In jeder Runde handlungssteuernde physiologische Belastungsanzeichen

Subjektive Belastungsreaktion

Wie bei der physiologischen werden auch die Indikatoren der subjektiven Belastungsfähigkeit in handlungsbegleitende und handlungssteuernde Belastungsanzeichen unterteilt. Sie betreffen die Dimensionen Emotion, Konzentration, Koordination und Wahrnehmung. Ein verlegenes Lachen oder die Zahleneingabe mit beiden Händen sind wahrnehmbare subjektive Belastungsanzeichen, haben jedoch keinen negativen Einfluss auf die Bearbeitung. Aus diesem Grund werden diese Beispielindikatoren der Kategorie handlungsbegleitende Belastungsanzeichen zugeordnet. Handlungssteuernde Indikatoren, die eine Bearbeitung negativ beeinflussen oder gar verhindern, sind in diesem Zusammenhang zum Beispiel Lachkrämpfe, Denkblockaden und Blackouts. Ebenfalls parallel zur vorherigen Subkategorie fungieren auch in Bezug auf die subjektive Befindlichkeit die Intensität der Verhaltensanker und die Anzahl der betroffenen Runden als Quantifizierungsregel.

Tabelle 4. Quantifizierung subjektive Belastungsreaktion

1	Keine subjektiven Belastungsanzeichen
2	In 1-2 Runden handlungsbegleitende subjektive Belastungsanzeichen (wahrnehmbar, Leistung wird nicht negativ beeinflusst)
3	In 3-5 Runden handlungsbegleitende subjektive Belastungsanzeichen
4	In jeder Runde handlungsbegleitende subjektive Belastungsanzeichen
5	In 1-2 Runden handlungssteuernde subjektive Belastungsanzeichen (so stark ausgeprägt, dass Leistung negativ beeinflusst)
6	In 3-5 Runden handlungssteuernde subjektive Belastungsanzeichen
7	In jeder Runde handlungssteuernde subjektive Belastungsanzeichen

4.1.3.2 Zuweisungsregeln Arbeitshaltung

Impulsivität vs. Reflexivität

Die Skala Impulsivität vs. Reflexivität soll beschreiben, mit welcher Geschwindigkeit und wie bedacht Entscheidungen im DCT-Szenario getroffen werden. Dabei stehen sich die Extremausprägungen einer schnellen und fehlerreichen sowie die der langsamen und fehlerarmen Bearbeitung gegenüber. Das Verhalten eines Bewerbers kann in einer Runde immer nur als impulsiv oder reflexiv beschrieben werden, je nachdem welche Verhaltensanker überwiegend beobachtet wurden. Je nach Art der Bearbeitung kann es also theoretisch vorkommen, dass auf dieser Skala eine einzige oder zwei verschiedene Noten vergeben werden. Die zweite Variante ist allerdings nur sehr selten der Fall. Können in allen Runden impulsive beziehungsweise reflexive Indikatoren beobachtet werden, ist dies mit der Note 7 zu bewerten. Den Schnittpunkt der beiden Extreme bildet eine schnelle und fehlerarme Vorgehensweise. Durch das bedachte Vorgehen bei einer hohen Bearbeitungsgeschwindigkeit kann die Problemaufgabe effektiv gelöst werden. Auf der Bewertungsskala ist dieser optimalen Ausprägung die Note 1 zugeordnet. Zuweisungskriterium für die Subkategorie Impulsivität vs. Reflexivität ist dementsprechend die Rundenanzahl, in denen die jeweiligen Verhaltensanker aus der Indikatorenliste überwiegend gezeigt wurden.

Tabelle 5. Quantifizierung Impulsivität vs. Reflexivität

7	Immer Impulsiv (<i>Bearbeitung schnell und fehlerreich, oberflächlich, Flüchtigkeitsfehler, Korrektur, Probieren, nicht Regelkonform</i>)
6	In 5 Runden vorwiegend impulsive Bearbeitung
5	In 4 Runden vorwiegend impulsive Bearbeitung
4	In 3 Runden vorwiegend impulsive Bearbeitung
3	In 2 Runden vorwiegend impulsive Bearbeitung
2	In 1 Runde vorwiegend impulsive Bearbeitung
1	In 6 Runden Effektiv (<i>Bearbeitung schnell und fehlerarm, Regelkonform</i>)
2	In 1 Runde vorwiegend reflexive Bearbeitung
3	In 2 Runden vorwiegend reflexive Bearbeitung
4	In 3 Runden vorwiegend reflexive Bearbeitung
5	In 4 Runden vorwiegend reflexive Bearbeitung
6	In 5 Runden vorwiegend reflexive Bearbeitung
7	Immer reflexive Bearbeitung (<i>langsam und fehlerarm, entscheidungsscheu, zögerlich</i>)

Frustrationstoleranz

Die beiden Extrempunkte dieser Skala setzten sich einerseits aus einer geringen, bewertet mit Note 7, andererseits aus einer hohen Toleranz von Rückschlägen zusammen. Als entscheidendes Zuweisungskriterium fungiert in dieser Subkategorie die Anzahl der gezeigten Verhaltensanker. Je nach Intensität wurden diese in starke und schwache Frustrationsanzeichen gegliedert. Ein deutlich wahrnehmbares Seufzen nach Rückschlägen ist ein Beispiel für schwach ausgeprägte Frustrationsanzeichen, wobei Fluchen und auf den Tischschlagen einer starken Beunruhigung entsprechen. Da es in jeder Runde zu zahlreichen frustrierenden Situationen kommen kann, diese jedoch je nach Bearbeitungsqualität stark variieren, bezieht man sich nicht auf die Anzahl der Runden, in denen die entsprechenden Indikatoren beobachtet wurden, sondern generell auf alle frustrierenden Situationen. Das Potential, Frustrationen hervorzurufen, haben alle Rückschläge, wie zum Beispiel

Überlastungen, Zeitmangel und Fehler des Partners. Für die Bewertung muss beurteilt werden, wie oft ein Bewerber Verhaltensanker einer starken beziehungsweise schwachen Frustrationstoleranz zeigt. Die Skala reicht hierbei von „nach Rückschlägen immer stark beunruhigt“ bis „nach Rückschlägen immer entspannt und motiviert“.

Tabelle 6. Quantifizierung Frustrationstoleranz

1	Hohe Frustrationstoleranz (nach Rückschlägen immer entspannt und motiviert)
2	Zeigt 1-2 mal schwache Frustration
3	Zeigt 3-5 mal schwache Frustration
4	Moderate Frustrationstoleranz - Zeigt 6 mal schwache Frustrationsanzeichen
5	Zeigt 1-2 mal starke Frustration
6	Zeigt 3-5 mal starke Frustration
7	Geringe Frustrationstoleranz (nach Rückschlägen immer stark beunruhigt)

Flexibilität

In der Subkategorie Flexibilität wird die Anpassungsfähigkeit an neue Situationen bewertet. Wie gelingt es, dem Bewerber sich von alten Verhaltensmustern zu lösen und seine Handlungen situativ zu optimieren. Der Umgang mit neuen Rollen und das Verhalten in unbekannten Situationen, wie fehlende Angaben, Systemausfällen sowie die unlösbaren Runden stehen dabei im Fokus der Bewertung. Zuweisungskriterium ist hier zum einen die Anzahl der Runden, in denen die Verhaltensanker vorwiegend gezeigt werden, und zum anderen die Intensität der Verhaltensanker. Dabei erfolgt eine Unterteilung in sofortige, verzögerte und keinerlei Anpassungsfähigkeit. Die optimale Ausprägung, eine sehr rasche Anpassungsfähigkeit in allen Runden, wird mit der Note 1 bewertet. Zeigt sich ein Bewerber in keinem der Teamdurchgänge flexibel, entspricht dies der Note 7.

Tabelle 7. Quantifizierung Flexibilität

1	Sehr flexibel – in jeder Runde flexibel, immer rasche Anpassung an neue Situationen und rollenkonform
2	In 5 Runden vorwiegend flexibel
3	In 4 Runden vorwiegend flexibel
4	Moderat flexibel - 3 x vorwiegend flexibel
5	In 2 Runden vorwiegend flexibel
6	In 1 Runden vorwiegend flexibel
7	Nicht flexibel – in 6 Runden nicht flexibel, keine Anpassung

4.1.3.3 Zuweisungsregeln Interaktion

Teamfähigkeit

Im DCT besteht eine Wechselbeziehung zwischen den beiden Akteuren. Es handelt sich um eine konjunktive Problemaufgabe, die nur von beiden Teammitgliedern gemeinsam gelöst werden kann (Stelling, 1998). Zur Zielerreichung müssen die beiden Bewerber also zusammenarbeiten. Die Qualität dieser Kooperation soll mit Hilfe der Subkategorie Teamfähigkeit bewertet werden. Je nachdem, in wie vielen Runden überwiegend Indikatoren der Zusammenarbeit in Abgrenzung zu denen der Einzelarbeit gezeigt werden, wird die Ausprägung der Teamfähigkeit bewertet. Die beiden Extrempunkte lauten dabei „*Einzelkämpfer*“ und „*sehr gute Kooperation*“.

Tabelle 8. Quantifizierung Teamfähigkeit

1	Sehr gut Kooperation (<i>immer gemeinsame Bearbeitung, Entscheidungsfindung, aktive Hilfe</i>)
2	In 5 Runden vorwiegend Kooperation
3	In 4 Runden vorwiegend Kooperation
4	Moderate Kooperation – in 3 Runden vorwiegend Kooperation
5	In 2 Runden vorwiegend Kooperation
6	In 1 Runde vorwiegend Kooperation
7	Keine Kooperation – (<i>Einzelkämpfer, keine Absprache, keine Hilfe</i>)

Dominanz vs. Submissivität

Die Subkategorie Dominanz vs. Submissivität beschreibt die soziale Kontrolle eines Bewerbers. Die Indikatoren dieser Kategorie wurden in dominante und submissive Verhaltensweisen unterteilt. So steht zum Beispiel die Fixierung auf die eigene Punktemaximierung dem selbstlosen Verzicht gegenüber. Je nach Anzahl der Runden, in welchen dominante oder submissive Verhaltensanker vorherrschen, erfolgt die entsprechende Notenvergabe. Die optimale Ausprägung wäre eine Bearbeitung auf Augenhöhe auf den Grundlagen sachlicher Diskussionen, respektvoller Umgangsformen und gemeinsamer Entscheidungen. Auch hier besteht die Möglichkeit der doppelten Notenvergabe in seltenen Ausnahmen.

Tabelle 9. Quantifizierung Dominanz vs. Submissivität

7	Dominanz – immer dominant (<i>Kontrolle, kein Unterordnen</i>)
6	In 5 Runden vorwiegend dominant
5	In 4 Runden vorwiegend dominant
4	In 3 Runden vorwiegend dominant
3	In 2 Runden vorwiegend dominant
2	In 1 Runden vorwiegend dominant
1	Gleichberechtigt (<i>mit Partner auf Augenhöhe, sachliche Diskussion, gemeinsame Entscheidungsfindung</i>)
2	In 1 Runden vorwiegend submissiv
3	In 2 Runden vorwiegend submissiv
4	In 3 Runden vorwiegend submissiv
5	In 4 Runden vorwiegend submissiv
6	In 5 Runden vorwiegend submissiv
7	Submissivität – immer submissiv (<i>Selbstlos, zurückhaltend</i>)

Kommunikation

Kommunikation ist, „was jemand von sich gibt beziehungsweise das, was beim anderen ankommt“ (Schulz von Thun, 1992, S.9). Die Güte der mündlichen Kommunikation im DCT resultiert aus den Fähigkeiten mündlicher Ausdruck,

Spracherkennung und Klarheit der Sprache. Auch die Informationsfilterung spielt in dieser Subkategorie eine bedeutende Rolle. Liest ein Bewerber alle Angaben ungekürzt vor oder ist er in der Lage, die wichtigsten Inhalte kurz und verständlich zu kommunizieren. Die Bewertung der kommunikativen Fähigkeit verläuft über die Rundenanzahl, in denen die Indikatoren einer problematischen Kommunikation gezeigt werden. Eine sehr gute Kommunikation, welche in allen Runden unproblematisch verläuft, wird mit der Note 1 bewertet. Der Bewerber drückt sich stets klar und verständlich aus, versteht die Aussagen des Partners und filtert immer die wichtigsten Informationen. Im Gegensatz dazu steht mit der Note 7 eine unzureichende Kommunikationsfähigkeit in allen sechs Durchläufen.

Tabelle 10. Quantifizierung Kommunikation

1	Sehr gute Kommunikationsfähigkeit – nie Kommunikationsprobleme
2	In 5 Runden keine Kommunikationsprobleme
3	In 4 Runden keine Kommunikationsprobleme
4	Moderate Kommunikationsfähigkeit –in 3 Runden keine Kommunikationsprobleme
5	In 2 Runden keine Kommunikationsprobleme
6	In 1 Runde keine Kommunikationsprobleme
7	Nicht ausreichende Kommunikationsfähigkeit - immer Kommunikationsprobleme

4.1.3.4 Zuweisungsregeln kognitive Leistung

Kognitive Leistung

Auch wenn der DCT in erster Linie nicht-kognitive Fähigkeiten erfassen soll, ist auch Kognition zur Lösung der Problemaufgabe unabdingbar. Wird diese Dimension außer Acht gelassen, wäre dies ein großer Informationsverlust, welcher die Qualität des Verfahrens mindern würde. Kognition bildet hierbei einen Sammelbegriff für den Vorgang und das Ergebnis der Informationsverarbeitung (Schuler, 1996). Im Zuge der Aufgabenbearbeitung müssen Regeln und Angaben gemerkt sowie Einschränkungen und Verhältnisse korrekt berechnet werden. Die zentralen kognitiven

Fähigkeiten lauten also Rechen- und Merkfähigkeit. Die Ausprägung dieser Fähigkeiten soll mit Hilfe der Subkategorie kognitive Leistung erhoben werden. Wichtige Verhaltensanker bilden hier die Umrechnungen von Verhältnissen und Einschränkungen sowie das Merken von Regeln, Angaben und Anweisungen. Ein Indikator für hohe kognitive Fähigkeiten sind freie Kapazitäten, welche eine Miteinbeziehung der verbleibenden Bearbeitungszeit und die Punktemaximierung ermöglichen. Zuweisungskriterium der kognitiven Leistungsfähigkeit ist die Rundenanzahl, in denen kognitive Probleme des Bewerbers wahrgenommen werden können. Zeigt ein Bewerber im Laufe der Bearbeitung freie Kapazitäten und verlaufen die Merk- und Rechenvorgänge in allen Runden fehlerfrei, entspricht dies einer sehr guten Bewertung mit der Note 1.

Tabelle 11. Quantifizierung kognitive Leistung

1	Sehr gute kognitive Leistung – keine kognitiven Probleme
2	In 1 Runde kognitive Probleme
3	In 2 Runde kognitive Probleme
4	Moderate kognitive Leistung - In 3 Runde kognitive Probleme
5	In 4 Runde kognitive Probleme
6	In 5 Runde kognitive Probleme
7	Ungenügende kognitive Leistung – immer kognitive Probleme

4.1.4 Bewertungsbogen DCT

Fügt man all diese Skalen zusammen, ergibt sich der folgende Bewertungsbogen (Abbildung 9). Er beinhaltet alle neun Skalen und jeweils ein Feld für den Vermerk „nicht beobachtet“. So werden keine Beurteilungen auf Grundlage unklarer Beobachtungen erzwungen. Solche Unklarheiten könnten zum Beispiel aus mangelnden Informationen oder nicht kategorisierbarer Verhaltensankern resultieren.

Belastbarkeit	Arbeitshaltung	Interaktion	Kognitive Leistung
<i>Physiologisch</i> ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	<i>Impulsivität vs. Reflexivität</i> 7 6 5 4 3 2 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	<i>Teamfähigkeit</i> ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7
<i>Subjektive Befindlichkeit</i> ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	<i>Frustrationstoleranz</i> 1 2 3 4 5 6 7 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	<i>Dominanz vs. Submissivität</i> 7 6 5 4 3 2 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	
	<i>Flexibilität</i> ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	<i>Kommunikation</i> ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7	

Abbildung 9. Beobachtungsbogen DCT

4.2 Beobachtertraining

Auf Basis der theoretischen Grundlagen wurde ein computergestütztes Beobachtertraining erstellt. Um die Schulungsinhalte an den bundeswehrinternen Rechnern auch offline durchführen zu können, wurde das Training mit dem Programm MS PowerPoint 2013 konstruiert. Neben der Kompatibilität bietet das Programm ebenfalls den Vorteil einer intuitiven Bedienung via Mausklick. Mit Hilfe von aktiven Flächen auf jeder Schulungsfolie wird ein dynamischer Aufbau ermöglicht. Der Anwender kann sich somit frei durch die Inhalte klicken und ganz gezielt bestimmte Kapitel aufrufen. Das Beobachtertraining untergliedert sich in die Abschnitte Einführung, Beobachterfehler, Dimensionen des DCT und einen abschließenden Übungsteil. Die Unterpunkte können als Äquivalent zu Kleinmanns (2003) Einteilung eines Beobachtertrainings in Beobachterfehlertraining, Dimensionsverwendungstraining, Bezugsrahmentraining und Verhaltensbeobachtungstraining gesehen werden.

Ziel des Beobachtertrainings ist es, fundiertes Wissen über die psychologische Verhaltensbeobachtung zu vermitteln. Theoretische Grundlagen, Definition und Ablauf einer wissenschaftlichen Verhaltensbeobachtung werden im ersten Kapitel aufgezeigt. Eine präzise Abgrenzung zu alltäglichen Beobachtungen sowie die Tragweite in Hinblick einer anschließenden Bewerberbeurteilung sollen ebenfalls verdeutlicht werden.

Die damit verbundenen Fehlerquellen werden anschließend im Abschnitt Beobachterfehler genauer beleuchtet. Dabei werden die im DCT-Szenario denkbaren Beobachter- und Beurteilungsfehler detailliert beschrieben und mit entsprechenden Praxisbeispielen sowie Karikaturen grafisch verdeutlicht. Die betreffenden Fehler lauten Halo-Effekt, Implizite Persönlichkeitstheorie, Primacy-/Recency-Fehler, Bezugsfehler, Similiar-to-me-Effekt, Grundeinstellung des Beobachters und Dominanz einzelner Beobachter. Im Anschluss an die Vorstellung werden für jeden Fehler gezielte Gegenmaßnahmen präsentiert.

Diese theoretischen Grundlagen sollen im nächsten Abschnitt in den Kontext des Dyadic Cooperation Tests und dessen Dimensionen übertragen werden. Bei diesem Abschnitt handelt es sich um ein Dimensionsverwendungstraining. Es werden die Konstrukte Belastbarkeit, Arbeitsstil, Interaktion und kognitive Leistung sowie die jeweiligen Unterkategorien und entsprechende Verhaltensanker genau erläutert. Für alle Verhaltensweisen werden explizite Zuordnungsregeln definiert, welche eine objektive Bewerberbeobachtung sicherstellen soll.

Um diese Zuordnungsregel zu festigen, folgt am Ende des Trainings ein Übungsteil. Dieser bildet eine Fusion aus Bezugsrahmentraining und Verhaltensbeobachtungstraining. Das gewonnene Wissen findet Anwendung mit Hilfe von praxisbezogenen Beispielen. Für diesen Zweck werden drei Videos von Bewerberdurchläufen gezeigt. Anschließend müssen in jeweils drei beispielbezogenen Fragen bestimmte Verhaltensweisen den entsprechenden Dimensionen zugeordnet und im nächsten Schritt bewertet werden. Nach jeder Antwort folgt ein Feedback bezüglich der Gültigkeit der gegebenen Antwort. Ist die Antwort falsch, muss eine neue Zuteilung erfolgen, bei richtigen Antworten erscheint die entsprechende Begründung.

Sind alle Übungsbeispiele bearbeitet, besteht die Möglichkeit, unklare Inhalte erneut aufzurufen und der Hinweis auf vertiefende Literatur.



Abbildung 10. Beispielfolien Beobachtertraining DCT

4.3 Methodik der Datenerhebung

Im Folgenden wird die Methodik der Datenerhebung zur Beantwortung der Fragestellung geschildert. Neben einer Beschreibung der Stichprobe werden das Untersuchungsdesign sowie die statistischen Auswertungsmethoden detailliert beschrieben.

4.3.1 Stichprobe

Zur Beantwortung der Fragestellungen wurden die Beobachtungsbögen von 33 DCT-Durchläufen analysiert. Im Erhebungszeitraum von Februar bis Mai 2014 wurde der Test von Bewerbern der militärischen Flugführungsdienste im Anschluss an das bisherige Auswahlverfahren am Zentrum für Luft- und Raumfahrtmedizin bearbeitet. Eine ungerade Bewerberanzahl hatte eine doppelte Bearbeitung Einzelner oder den Einsatz bundeswehrinterner Mitarbeiter als Dummies zur Folge. Die DCT-Bearbeitung dieser Personen wurde von der Beobachterkommission nicht bewertet. Dieses Vorgehen erklärt, dass in den 33 Durchläufen 63 Personen bezüglich der neun Subkategorien benotet wurden. Die Stichprobe setzt sich aus 7 weiblichen und 56 männlichen Testpersonen zusammen. Die konträre Geschlechtsverteilung erklärt sich durch die Überpräsenz männlicher Bewerber der militärischen Flugführungsdienste bei der Bundeswehr. Die Bewertung erfolgte paarweise von sechs variierenden

Diplompsychologen und einem militärischen Prüffizier, welcher gleichzeitig als Testleiter fungierte. Unter den Psychologen befanden sich zwei Frauen.

4.3.2 Untersuchungsdesign

Das Untersuchungsdesign unterteilt sich in zwei Messzeitpunkte. Der erste Messzeitpunkt vollzog sich von 26.02. bis 30.04.2014. In dieser Zeit durchliefen 31 Bewerber das Auswahlverfahren am Zentrum für Luft- und Raumfahrtmedizin und bearbeiteten in diesem Zusammenhang das DCT-Verfahren. Ab 07.04.2014 wurden die Psychologen und der Prüfstaboffizier staffelweise mit Hilfe des computer-gestützten Trainingsprogramms in Bezug auf die wissenschaftliche Verhaltensbeobachtung des DCTs geschult. Dieser Vorgang erstreckte sich bis 02.05.2014. Die ersten DCT-Durchgänge, welche von geschulten Beobachtern bewertet wurden, starteten am 30.04.2014. Der zweite Messzeitpunkt endete mit den finalen Erhebungen am 16.05.2014. Es resultieren jeweils 63 Bewertungen der neun Skalen des DCTs. Das Untersuchungsdesign wird in folgender Abbildung nochmals schematisch verdeutlicht.

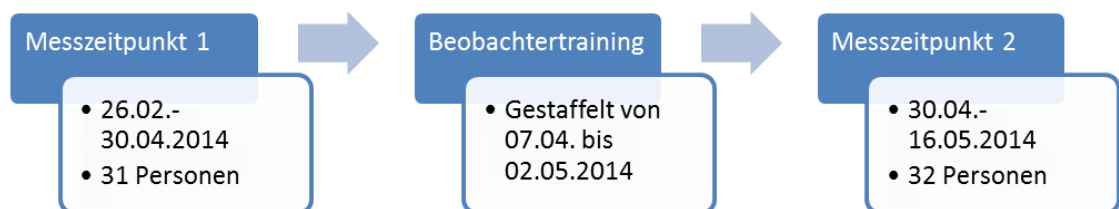


Abbildung 11. Untersuchungsdesign

4.3.3 Datenaufbereitung

Die Beobachtungsbögen wurden im Anschluss an den zweiten Messzeitpunkt digitalisiert. Die entsprechenden Variablen wurden der Papierform entnommen und via SPSS 14 aufbereitet. Sowohl die deskriptive Betrachtung als auch die statistische Analyse der Daten erfolgte mit Hilfe der computergestützten Programme MS Excel 2010 und SPSS 14. Hintergründe der verwendeten Auswertungsverfahren und das genaue Vorgehen werden in den folgenden Abschnitten im Detail erläutert.

4.3.3.1 Deskriptiv

Die von den Beobachter abgegebenen Bewertung bezüglich der Skalen *physische und subjektive Belastbarkeit, Impulsivität vs. Reflexivität, Frustrationstoleranz, Flexibilität, Teamfähigkeit, Dominanz, Kommunikation* sowie *kognitive Leistung* werden deskriptiv dargestellt. Dies beinhaltet Häufigkeiten, Mittelwerte und Standardabweichungen. Die Inhalte werden gelistet und im Sinne der Anschaulichkeit auch grafisch aufbereitet. Dabei werden einerseits der Gesamtdatensatz und andererseits die beiden Messzeitpunkte getrennt betrachtet.

4.3.3.2 Beobachterübereinstimmung

Nach Ausarbeitung der Skalen und Erstellung des Beobachtertrainings gilt es, die Güte der vorhergegangenen Arbeitsschritte zu prüfen. Wie im oberen Teil der Diplomarbeit bereits angeführt wurde, bilden Reliabilität, Validität und Objektivität die zu prüfenden Gütekriterien.

Die Wiederdurchführung des Verfahrens, zur Berechnung der Retest-Reliabilität ist aufgrund von Erinnerungs- und Übungseffekten in diesem Zusammenhang nicht möglich. Auch die prognostische Validität kann erst nach fortgeschrittener Ausbildung der Bewerber erhoben werden. Sie ist somit ebenfalls kein Gegenstand dieser Studie. Entsprechende Analysen folgen zu einem späteren Zeitpunkt mit den Daten aus der Langzeiterfolgskontrolle (LEK). Die Konstruktvalidität bildet einen kleinen Anteil der Gesamtvalidität. Diese kann mit Hilfe von Expertenratings und wissenschaftlicher Literatur erzielt werden (Kelava & Moosbrugger, 2008, S. 15-18). Derartige Validitätsstudien sollten ebenfalls an diese Diplomarbeit anknüpfen.

Zur Überprüfung der Verfahrensgüte dient im Rahmen dieser Arbeit die Objektivität. Ein Maß für die Objektivität ist die Interrater-Reliabilität, also das Ausmaß der Übereinstimmung verschiedener Beobachter. Hierfür können Ähnlichkeit- und Übereinstimmung der Beobachtung berechnet werden. Hohe Übereinstimmungen von Beurteilungen sprechen für eine Zuverlässigkeit von Einzelurteilen. Beobachterübereinstimmungen können mit Hilfe von Prozentangaben oder grafischen Darstellungen veranschaulicht werden. Die Berechnung der Korrelationskoeffizienten hat in der Forschung jedoch eine herausragende

Bedeutung, da sie in Hinblick auf die Klassische Testtheorie Rückschlüsse auf die Reliabilität zulassen (Rindermann, 2002).

Einen ausführlichen Überblick geben Caspar & Wirtz (2002). Im vorliegenden Untersuchungsdesign wurden die Personen von zwei bis drei Ratern mit Hilfe einer absoluten Ratingskala beurteilt. Dabei bewertete nicht jeder Rater jede Versuchsperson. Man spricht also von einer rater-randomisierten Vorgehensweise. In Hinblick auf diese Besonderheit des Datensatzes kann die Reliabilität mit Hilfe der unjustierten Intraklassenkorrelation ($ICC_{unjustiert}$) erfasst werden. Bei der Berechnung wird die relative Lage der Ratings zueinander verglichen. Die Voraussetzungen für die Berechnung gleichen einer einfaktoriellen ANOVA. Neben intervallskalierten Daten müssen Normalverteilung und Varianzhomogenität gegeben sein. Das Zusammenhangsmaß hat einen Wertebereich von -1 bis +1, wobei +1 theoretisch einem perfekten Beobachterzusammenhang entspricht. Da eine negative Reliabilität nicht existiert, werden alle negativen Werte mit Ausprägung 0 interpretiert. Bei der Berechnung resultieren zwei verschiedene Korrelationskoeffizienten. Die $ICC_{unjustiert}$ Single Measure bezieht jedes einzelne Rating in die Korrelation ein. Hingegen setzt die $ICC_{unjustiert}$ Average Measure bei den Mittelwerten der Rater an. Aus diesem Grund fällt der zweite Koeffizient meist höher aus. Die Signifikanz spielt hierbei eine untergeordnete Rolle und wird deshalb auch nur selten getestet. Die Fragestellung, ob die Übereinstimmung besser ist als es der pure Zufall erwarten lässt, ist nicht entscheidend. Da sich die viel wichtigeren Informationen aus den absoluten Ausprägungen der Koeffizienten ableiten lassen, stehen diese im Interessenfokus dieser Arbeit. Im Kontext der ICC hat also das Erreichen einer bestimmten Mindestausprägung einen viel höher Stellenwert. Ab einen Wert von 0,4 spricht man von akzeptablen Resultaten. Für eine gute Reliabilität sollten die Korrelationsmaße eine Mindestausprägung von 0,7 haben (Caspar & Wirtz, 2002). Die Standardmatrix zu Berechnung der ICC unterliegt folgendem Aufbau. Die Bewertungen von k Ratern in Bezug auf n Personen werden wie in Abbildung 12 gelistet.

Tabelle 12. Standardmatrix zur Berechnung der ICC (nach von Eye und Mun, 2005)

<i>Case(i)</i>	<i>Rater(j)</i>					
	1	2	3	*	*	k
1	$y_{1.1}$	$y_{1.2}$	$y_{1.3}$	*	*	$y_{1.k}$
2	$y_{2.1}$	$y_{2.2}$	$y_{2.3}$	*	*	$y_{2.k}$
3	$y_{3.1}$	$y_{3.2}$	$y_{3.3}$	*	*	$y_{3.k}$
*	*	*	*	*	*	*
*	*	*	*	*	*	*
n	$y_{n.1}$	$y_{n.2}$	$y_{n.3}$	*	*	$y_{n.k}$

In dieser Studie werden die 63 Bewerber auf neun verschiedenen Dimensionen bewertet. Die Beurteilung eines Bewerbers erfolgt parallel von zwei bis drei Ratern. Zur Berechnung der ICC werden alle Ratings einbezogen. Die folgende Tabelle zeigt die verwendete Datenmatrix zur Berechnung der ICC.

Tabelle 13. Datenmatrix zur Berechnung der ICC

<i>Case(i)</i>	<i>Rater(j)</i>		
	1	2	3
Phy_1	$y_{1.1}$	$y_{1.2}$	$y_{1.3}$
Sub_1	$y_{2.1}$	$y_{2.2}$	$y_{2.3}$
Imp_1	$y_{3.1}$	$y_{3.2}$	$y_{3.3}$
FT_1	$y_{4.1}$	$y_{4.2}$	$y_{4.3}$
Flex_1	$y_{5.1}$	$y_{5.2}$	$y_{5.3}$
Team_1	$y_{6.1}$	$y_{6.2}$	$y_{6.3}$
Dom_1	$y_{7.1}$	$y_{7.2}$	$y_{7.3}$
Kom_1	$y_{8.1}$	$y_{8.2}$	$y_{8.3}$
Kog_1	$y_{9.1}$	$y_{9.2}$	$y_{9.3}$
Phy_2	$y_{10.1}$	$y_{10.2}$	$y_{10.3}$
*	*	*	*
Kog_63	$y_{567.1}$	$y_{567.2}$	$y_{567.3}$

4.3.3.3 Beobachterfehler

Für den Befund möglicher Beobachterfehler wird ein besonderes Augenmerk auf die Präsenz von zentralen Tendenzen, Milde- und Strengeneffekten gelegt. Diese Betrachtung wird vor und nach dem Beobachtertraining durchgeführt. Ein Strengeneffekt liegt vor, wenn die Bewertungen eines Beobachters mehr als einen Skalenwert in negativer Richtung von der durchschnittlichen Bewertung abweicht. Ist die Differenz zwischen Einzelwertung und Durchschnitt größer als ein Skalenwert

in positiver Richtung, liegt ein Mildefehler vor (Eschen-Léguedé, 2012, S.51). Zentrale Tendenzen zeigen sich durch Meidung der Extremnotenvergabe. In Bezug auf das Vorhandensein von zentralen Tendenzen, Milde und Strengeeffekten werden die beiden Zeitpunkte vor und nach dem Beobachtertraining miteinander verglichen. Dies geschieht durch die Betrachtung der Häufigkeitsverteilungen.

5. ERGEBNISSE

Dieses Kapitel zeigt, auf welchem Weg die Daten des DCT-Ratings aufbereitet und ausgewertet wurden. Außerdem werden die Fragestellungen der Forschungshypothesen mit statistischen Analysen im Detail untersucht. Demzufolge orientiert sich die Gliederung des Ergebnisteils an der Aufbereitung der in Kapitel 3 vorgestellten Forschungsfragen sowie den entsprechenden Hypothesen. Die resultierenden Ergebnisse werden im darauf folgenden Kapitel 6 umfassend diskutiert.

5.1 Deskriptiv

Die deskriptive Betrachtung soll einen ersten Überblick über die vorliegenden Daten geben, auf welchen die anschließenden Analysen basieren. Es werden Anzahl, Mittelwerte und Standardabweichung der Notenverteilung differenziert für jede der Dimensionen einzeln angeführt. Hierfür wurden die Beurteilungen der einzelnen Beobachter aufsummiert. Im Fokus der Betrachtung stehen sowohl der Gesamtdatensatz als auch die getrennte Betrachtung der beiden Messzeitpunkte. Die Häufigkeiten pro Kategorie sind mit absoluten Zahlen angeführt und werden in der untenstehenden Abbildung visualisiert.

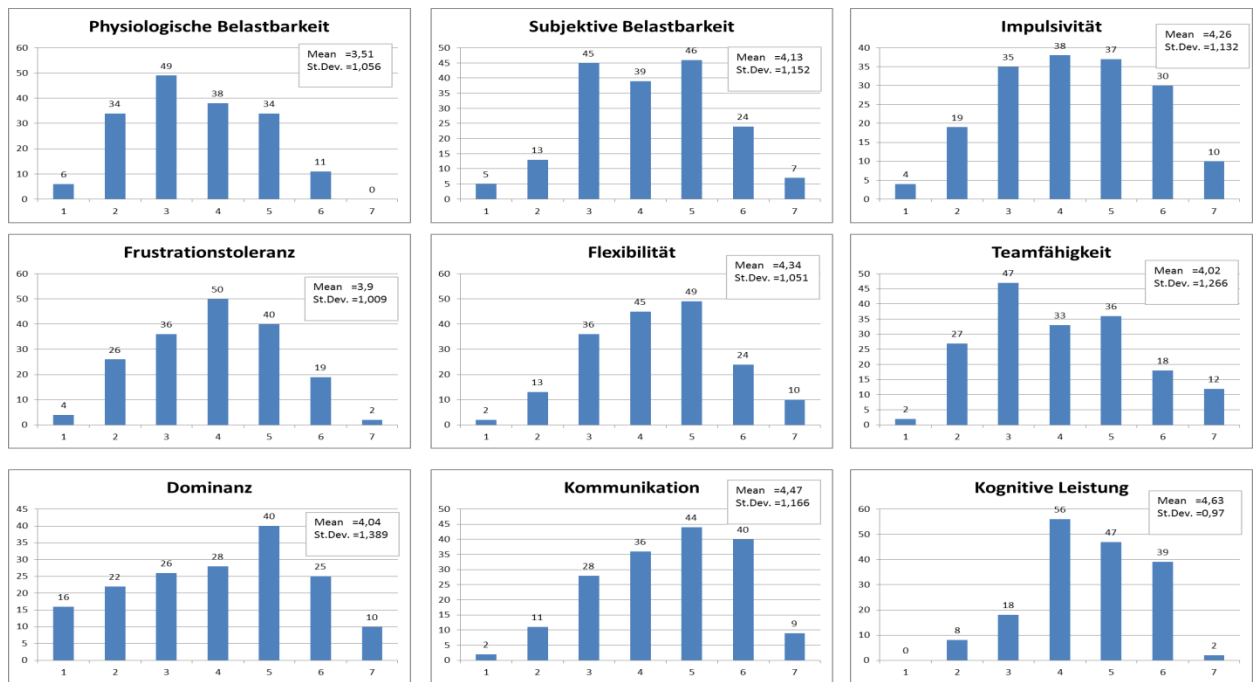


Abbildung 12. Häufigkeitsverteilung der bewerteten Skalen gesamt

Die grafische Betrachtung zeigt, dass die Notenvergabe bei den meisten Dimensionen annähernd normalverteilt ist. Die Normalverteilung zeichnet sich durch die gehäufte Notenvergabe im durchschnittlichen Bereich und ein kontinuierliches Absinken der Häufigkeit bis hin zu den Randnoten 1 und 7 aus. Ausnahmen bilden hierbei die Skalen *physiologische Belastbarkeit* mit einer rechtsschiefen sowie *Kommunikation* und *kognitive Leistung* mit einer linksschiefen Verteilung. Die Mittelwerte dieser Skalen weichen auf die erste Kommastelle gerundet um mindestens 0,5 Skalenwerte von der Durchschnittsnote 4 ab. Des Weiteren fällt auf, dass die Skala *Dominanz* mit 1,389 eine erhöhte Standardabweichung besitzt. Die durchschnittliche Entfernung vom Mittelwert beträgt hier 1,389 Notenpunkte. Es resultiert eine weniger differenzierte Verteilung. Die Gesamtdurchschnittsnote über alle Dimensionen hinweg beträgt 4,2.

Betrachtet man die Abbildungen 13 und 14, also die einzelnen Skalen getrennt für beide Messzeitpunkte, zeichnet sich eine ähnliche Verteilung der Skalen ab. In Bezug auf den Erhebungszeitraum vor dem Beobachtertraining fällt auf, dass die Randkategorien nur sehr selten und zum Teil gar nicht besetzt sind. Die durchschnittliche Bewertung aller Skalen liegt zum ersten Messzeitpunkt bei 4,0.

Nach dem Beobachtertraining erhöht sich dieser Wert auf 4,3. Durch einen Blick auf Abbildung 14 wird deutlich, dass die Standardabweichung der Skala *Dominanz* nach dem Beobachtertraining mit einem Wert von 1,693 höher als vor dem Training ist. Auch in Bezug auf die Skala *Teamfähigkeit* steigt die Standardabweichung von 1,022 auf 1,491. Es resultiert eine undifferenzierte Verteilung. Außerdem ist festzustellen, dass es zum zweiten Messzeitpunkt öfter zu den Randbewertungen 1 und 7 kommt.

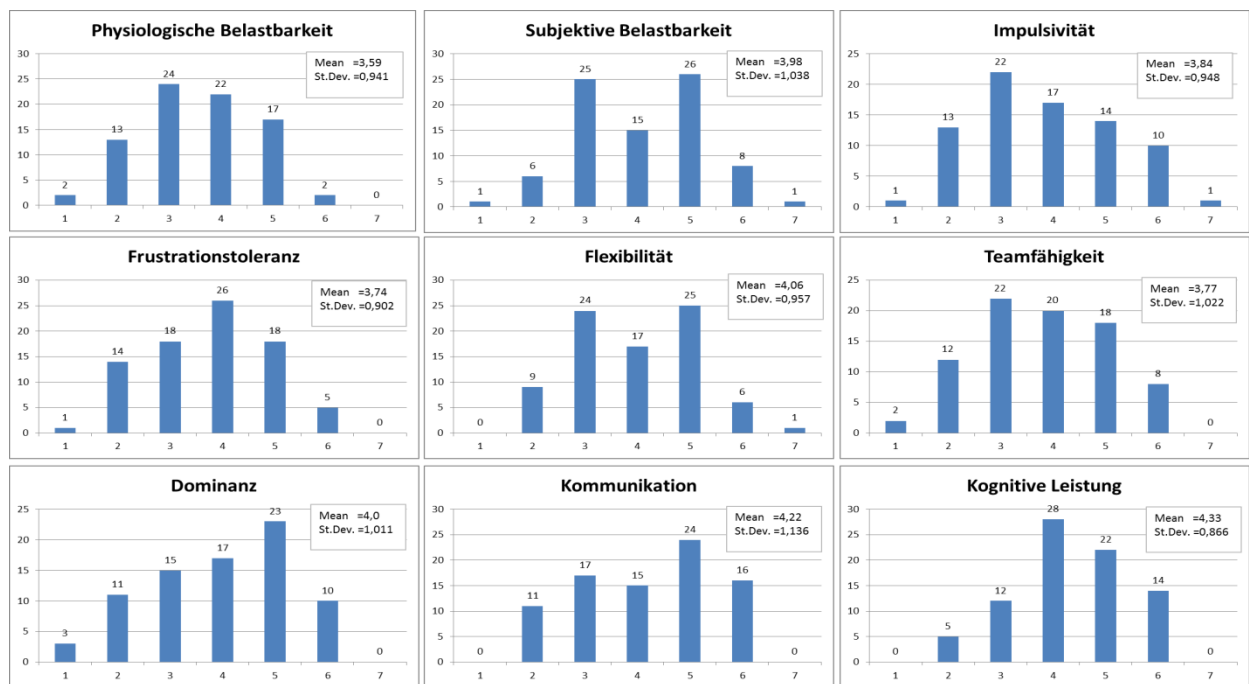


Abbildung 13. Häufigkeitsverteilung der bewerteten Skalen vor dem Training

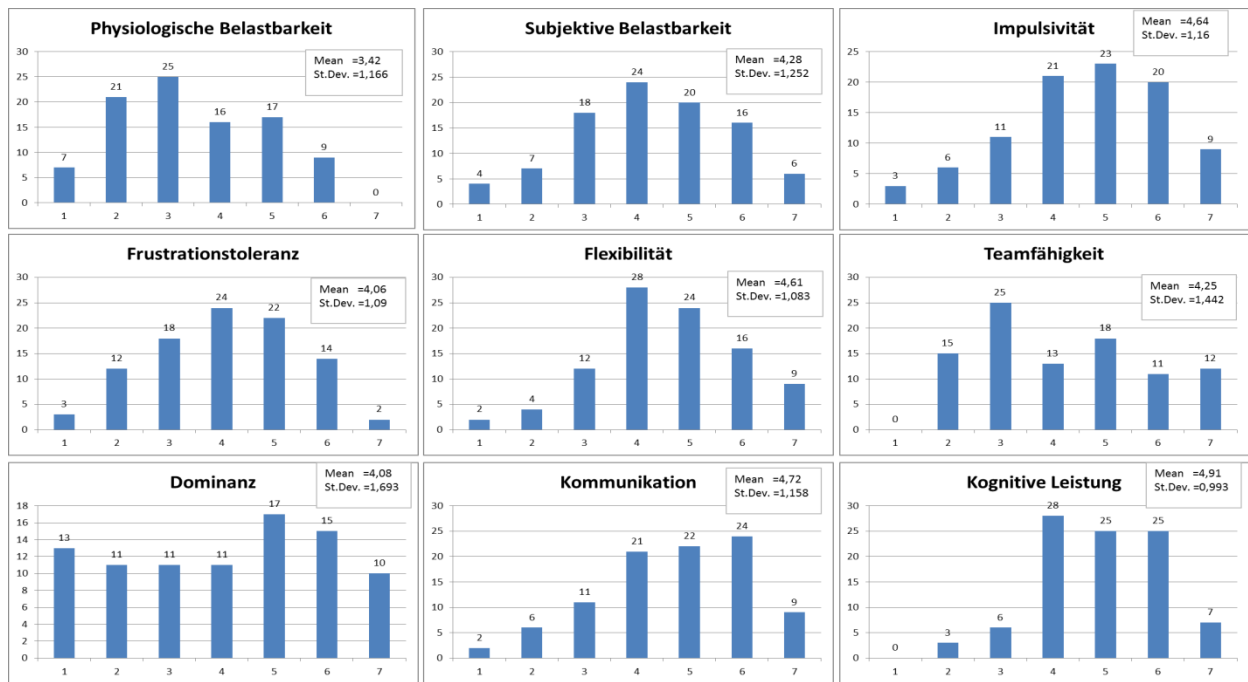


Abbildung 14. Häufigkeitsverteilung der bewerteten Skalen nach dem Training

Die Nutzung der Antwortalternative „*nicht beobachtet*“ entspricht den sogenannten Missings im Datensatz. Betrachtet man die entsprechenden Häufigkeiten, zeichnet sich folgendes Bild ab. Insgesamt konnten die entsprechenden Kategorien 19-mal nicht bewertet werden, weshalb von den Beobachtern das Feld „*nicht beobachtet*“ markiert wurde. Vor dem Beobachtertraining trat dieser Fall 11-mal ein. Besonders häufig wurde die Antwortalternative in Bezug auf die Dimensionen *Impulsivität* (4 Missings) und *Dominanz* (3 Missings) genutzt. Nach Durchführung des Beobachtertrainings konnte die Anzahl der Missings auf 8 reduziert werden. Die größte Anzahl ausstehender Bewertungen hat auch zum zweiten Messzeitpunkt die Dimension *Dominanz*, mit 4 Missings. Die Nutzung der Alternative „*nicht beobachtet*“ hat sich jedoch ausschließlich in diesem Bereich leicht erhöht und ist insgesamt gesunken.

5.2 Beobachterübereinstimmung

Mit dem Fokus auf die einzelnen Beobachter zeigen die Daten in der unten angeführten Tabelle 15, dass die Anzahl der jeweiligen Beurteilungen stark variiert. So hat der Beobachter 1 vor der Schulung 23 und nach der Schulung 29 Bewerber bewertet. Hingegen hat Beobachter 7 nur einen Bewerber vor dem Training bewertet. Nach Durchlaufen des Beobachtertrainings hat Beobachter 7 gar keine Bewertungen abgegeben. Diese Tatsache führt zu einer löchrigen Datenmatrix mit einer hohen Anzahl an fehlenden Bewertungen. Aus diesem Grund wird in den folgenden Analysen auf die Aufgliederung in die sieben Beobachter verzichtet. Die Bewertungen werden für jede Beobachtung zusammengefasst. Es resultieren die Ratings der unspezifischen Beobachter 1 und 2 sowie in 52 Fällen die eines dritten Beobachters, losgelöst von der jeweiligen Rateridentität. Diese variierende Personenzahl liegt in den Dienstplänen der Mitarbeiter des Standortes begründet. Aufgrund von diversen Dienstreisen und differierender Abwesenheit stand nicht an allen Erhebungstagen die gewünschte Anzahl, von drei Beurteilern zur Verfügung.

Tabelle 14. Personenspezifische Bewertungsanzahl vor und nach Schulung

Beobachter	1	2	3	4	5	6	7
Vor	23	15	11	10	2	9	1
Schulung							
Nach	29	12	5	15	3	24	0
Schulung							

Das Ausmaß der Beobachterübereinstimmung wird mit Hilfe der unjustierten Intraclass Correlation ($ICC_{\text{unjustiert}}$) erfasst. Die Voraussetzungen für die Berechnung entsprechen einer einfaktoriellen Varianzanalyse (ANOVA). Die Daten der 7-stufigen Ratingskala sind intervallskaliert. Eine Normalverteilung der Variablen als Voraussetzung für die ANOVA liegt nicht in allen Fällen vor. Diese Tatsache zeichnet sich schon grafisch in den Abbildungen der deskriptiven Betrachtung ab. Da es sich bei diesen Darstellungen um die aufsummierten Häufigkeiten der Einzelratings handelt, wurde notwendigerweise auch die Normalverteilung der neun

Skalen raterspezifisch geprüft. Die rechnerische Testung unter Verwendung des Kolmogorov Smirnov Verfahrens (KS-Test) konnte zeigen, dass die Variablen *Physiologische Belastungsfähigkeit*, *Flexibilität*, *Teamfähigkeit* und *Kognitive Leistung* von Beobachter 1, sowie *physiologische Belastbarkeit*, *subjektive Belastbarkeit*, *Teamfähigkeit* und *kognitive Leistung* bei Beobachter 2 nicht normalverteilt sind. Dies trifft auch auf die Variablen *Kommunikation* und *kognitive Leistung* des 3. Beobachters zu. Obwohl dies 10 der 27 Variablen entspricht, kann man jedoch davon ausgehen, dass die ANOVA und somit auch die ICC_{unjustiert} robust auf die Verletzung der Normalverteilung reagieren (Caspar & Wirtz, 2002, S.180).

Eine weitere Voraussetzung für die Berechnung einer ANOVA ist die Homogenität der Varianzen. Sie wird unter Zuhilfenahme des Levene Tests berechnet. Da bei der Prüfung der Voraussetzungen die Nullhypothese der Wunschhypothese entspricht, setzt das verwendete Signifikanzniveau bei $\alpha = 0,25$ an (Caspar & Wirtz, 2002, S.180). Die Varianzhomogenität liegt nur für eine der 27 Variablen, nämlich bei *Teamfähigkeit Beobachter 2* nicht vor. Durch die Verletzung der Voraussetzung dürfen die Werte nicht interpretiert werden. Demzufolge muss die Skala *Teamfähigkeit* von der folgenden Betrachtung ausgeschlossen werden. Für die restlichen Skalen kann die Varianzhomogenität jedoch als gegeben angesehen werden. Folglich darf die ICC_{unjustiert} als Maß für die Beobachterübereinstimmung berechnet werden. Die Ergebnisse werden in der unteren Tabelle angeführt.

Tabelle 15 . Die Intraclass-Correlation (ICC) der einzelnen Skalen, Vertrauensintervalle in Klammern

Dimension	Gesamt		Vor Schulung		nach Schulung	
	Single	Average	Single	Average	Single	Average
<i>Physische Belastbarkeit</i>	,486 [,317; ,643]	,739 [,582; ,844]	,593 [,329; ,803]	,814 [,596; ,924]	,452 [,235; ,656]	,712 [,48; ,851]
<i>Subjektive Belastbarkeit</i>	,538 [,379; ,682]	,778 [,647; ,865]	,496 [,23; ,733]	,747 [,472; ,892]	,559 [,356; ,734]	,792 [,624; ,892]
<i>Impulsivität</i>	,588 [,428; ,727]	,810 [,692; ,889]	,305 [,006; ,629]	,569 [,019; ,836]	,698 [,527; ,829]	,874 [,77; ,936]
<i>Frustrations- toleranz</i>	,341 [,168; ,518]	,608 [,377; ,763]	,243 [-,021; ,544]	,49 [-,064; ,781]	,381 [,16; ,6]	,648 [,364; ,818]
<i>Flexibilität</i>	,413 [,242; ,58]	,679 [,49; ,806]	,22 [-,04; ,525]	,459 [-,129; ,768]	,407 [,188; ,621]	,673 [,41; ,831]
<i>Teamfähigkeit</i>	-	-	-	-	-	-
<i>Dominanz</i>	,601 [,436; ,743]	,819 [,698; ,896]	,423 [,128; ,704]	,687 [,306; ,877]	,697 [,507; ,839]	,873 [,755; ,94]
<i>Kommunikation</i>	,453 [,283; ,615]	,713 [,542; ,828]	,445 [,167; ,705]	,707 [,376; ,877]	,431 [,213; ,64]	,694 [,447; ,842]
<i>Kognitive Leistung</i>	,517 [,352; ,668]	,763 [,62; ,858]	,503 [,218; ,736]	,744 [,455; ,893]	,516 [,302; ,706]	,762 [,565; ,878]

Betrachtet man die Resultate der ICC_{unjustiert} über beide Messzeitpunkte hinweg, ergibt sich folgendes Bild. Der Wertebereich der Skalen liegt bezüglich der Single Measure bei 0,341 bis 0,601. In Anlehnung an die Einteilung von Caspar und Wirtz (2002) können diese Werte wie folgt interpretiert werden. Mit Ausnahme der Skala *Frustrationstoleranz* zeigen alle Skalen Werte größer als 0,4 und somit akzeptable Interrater-Reliabilitäten. Die Average Measure vergleicht die Ratermittelwerte und erzielt generell höhere Werte als die Single Measure. Die Resultate befinden sich in einem Bereich von 0,608 bis 0,819. Dabei sprechen nur zwei der betrachteten Skalen, *Frustrationstoleranz* und *Flexibilität* mit einem Wert unter 0,7 für eine

akzeptable Beobachterübereinstimmung. Die Ergebnisse der restlichen Skalen sprechen bei dieser Berechnungsweise für gute Reliabilitäten.

Die Ergebnisse der Single Measures vor dem Training befinden sich in einem Wertebereich von 0,220 bis 0,593. Dies entspricht nach der Einteilung von Caspar und Wirtz (2002) nicht akzeptabler ($<0,4$) bis akzeptabler ($<0,7$) Ausprägungen. Nach dem Training verschiebt sich der Wertebereich der Single Measure auf 0,381 bis 0,698. Somit ist die Beobachterreliabilität nur noch in Bezug auf die Skala *Frustrationstoleranz* nicht akzeptabel. Alle anderen Skalen befinden sich in einem akzeptablen Bereich. Auch in Hinblick auf die Average Measure, die im Allgemeinen zu höheren Ergebnissen führt, zeichnet sich ein ähnliches Bild ab. Der Wertebereich liegt vor dem Training bei 0,459 bis 0,814. Dabei überschreiten die Skalen *Physische Belastbarkeit*, *Subjektive Belastbarkeit*, *Kommunikation* und *kognitive Leistung* den Grenzwert von 0,7, was für eine gute Interrater-Reliabilität spricht. Nach dem Training verfehlt *Kommunikation* mit einem Wert von 0,687 nur knapp das Prädikat gut und liegt im akzeptablen Bereich. Es sprechen nun auch die Werte der Skalen *Dominanz* (0,873) und *Impulsivität* (0,874) für eine gute Beobachterreliabilität. In diesen beiden Bereichen zeigt sich der Effekt des Beobachtertrainings auf die Beobachterreliabilität am deutlichsten.

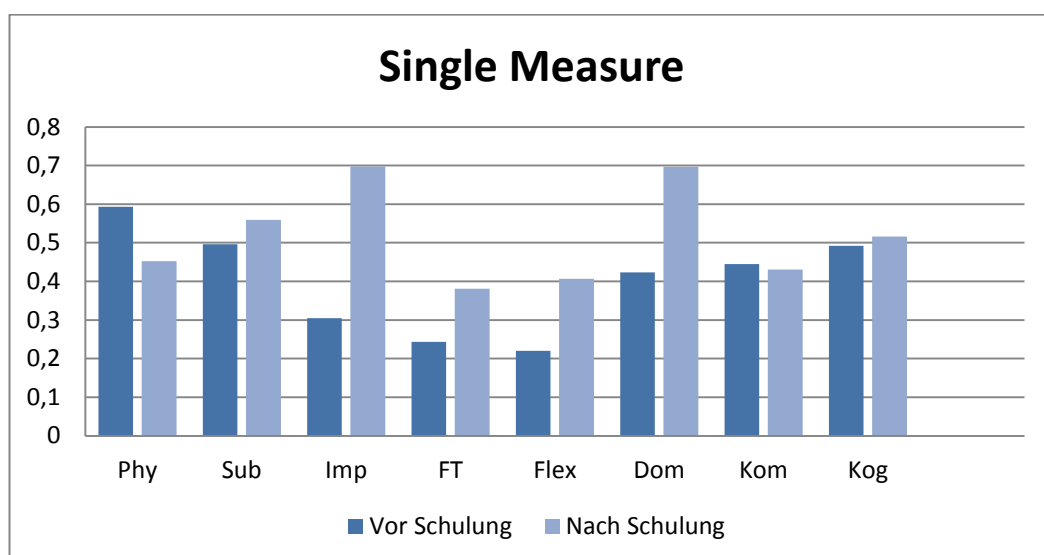


Abbildung 15. Single Measure vor und nach Beobachterschulung

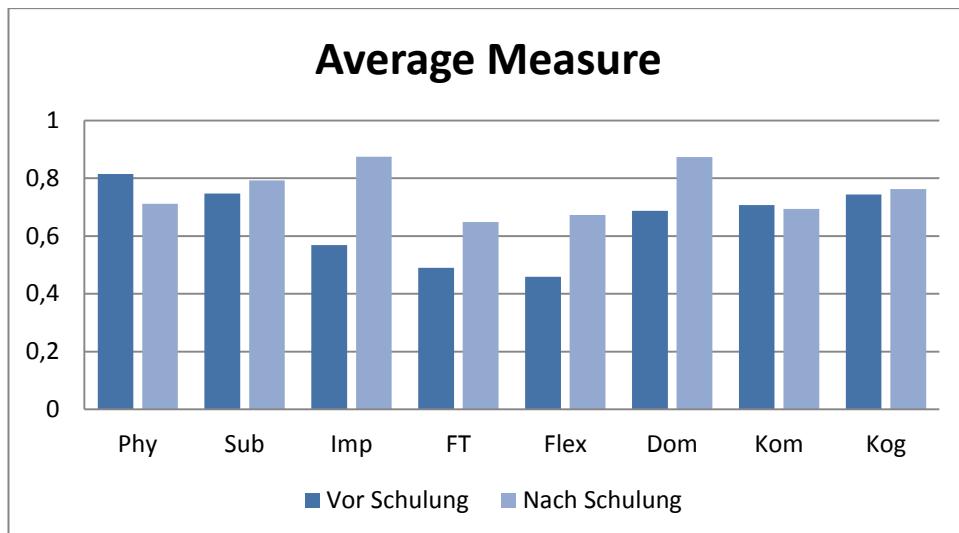


Abbildung 16. Average Measure vor und nach Beobachterschulung

Betrachtet man die oberen Abbildungen 15 und 16, die Werte der $ICC_{unjustiert}$ in Bezug auf die einzelnen Skalen, fällt auf, dass es nach dem Beobachtertraining in fast allen Bereichen zu einer Verbesserung der Beobachterübereinstimmung gekommen ist. Die einzigen beiden Ausnahmen bilden die Skalen *physiologische Belastbarkeit* und *Kommunikation*. Hier hat die Einigkeit der Beurteiler sowohl in Bezug auf die Single Measure als auch auf die Average Measure leicht abgenommen.

In Tabelle 15 sind hinter den Ergebnissen der $ICC_{unjustiert}$ die jeweiligen Vertrauensintervalle in Klammern angeführt. Der wahre Wert der Person liegt mit 95 prozentiger Sicherheit innerhalb dieses Vertrauensintervalls. Betrachtet man die vorliegenden Vertrauensintervalle wird deutlich, dass der Wertebereich beider Berechnungen nach dem Training geringer wird. Dies trifft auf alle Skalen zu. Man geht davon aus, dass eine höhere Reliabilität mit einem geringeren Standardmessfehler und somit mit kleineren Vertrauensintervallen einhergeht.

5.3 Beobachterfehler

Vor der Schulung der Beobachter liegt der Mittelwert aller Urteile bei 4,0 und einer Standardabweichung von 0,72. Zum zweiten Messzeitpunkt sind die Bewertungen im Durchschnitt etwas strenger. Der gemittelte Wert lautet 4,3, bei einer Standardabweichung von 0,91. Tabelle 16 zeigt die gemittelten Einzelbewertungen

und ihre Differenz zur durchschnittlichen Bewertung aller Beobachter. Vergleicht man diese Werte fällt auf, dass die mittleren Einzelratings innerhalb und zwischen den beiden Messzeitpunkten durchaus variieren. Die Ratings der Beobachter 1 und 3 liegen immer etwas unter dem Gesamtdurchschnitt und sind somit vergleichsweise milde. Rater 4, 6 und 7 urteilen etwas strenger als der Durchschnitt aller Beobachter. Die Differenz ist jedoch bei keinem Beobachter und zu keinem Zeitpunkt größer als ein Skalenwert. Die Beobachtungsfehler Tendenz zur Milde und Tendenz zur Strenge konnten weder vor noch nach der Beobachterschulung beobachtet werden.

Tabelle 16. Differenz Mittel Einzelbewertung - Mittel Gesamtbewertung

Beobachter		1	2	3	4	5	6	7
Vor Schulung $\bar{x} = 4,0$	Mittelwert	3,92	3,76	3,57	4,15	4,04	4,3	4,3
	Differenz	0,08	0,24	0,43	-0,15	- 0,04	- 0,3	- 0,3
	Mittelwert	4,13	4,7	3,71	4,5	4,17	4,43	-
	Differenz	0,17	-0,4	0,59	- 0,2	0,13	-0,13	-

Die Berechnungen zeigen, dass die durchschnittliche Bewertung nach der Schulung um 0,3 Skalenwerte strikter geworden ist. Nun stellt sich die Frage, ob die Rater nach dem Training tatsächlich strenger in ihrer Beurteilung geworden sind oder ob die Mittelwertsunterschiede durch einen Leistungsunterschied der beiden Bewerbergruppen zu erklären sind. Für die Beantwortung dieser Fragestellung wurde der in den Einzelrunden erzielte Solowert³ herangezogen.

³ Wie bereits in Abschnitt 1.2 beschrieben, werden vor der Teamaufgabe vier Einzelrunden bearbeitet, um so eine Baseline der Bewerberfähigkeit zu gewinnen. Als Vergleichseinheit fungiert der in den Solorunden erspielte Punktwert.

Ein Vergleich der entsprechenden Mittelwerte mittels T-Test für unabhängige Stichproben zeigt bei einem Signifikanzniveau von 0,5 einen signifikanten Leistungsunterschied der beiden Messzeitpunkte. Der t-Wert beträgt 4,065 und $p \leq 0,001$. Die Bewerbergruppe vor dem Beobachtertraining konnte mit einem durchschnittlichen Solowert von 311,78 Punkten deutlich besser abschneiden als die Bewerbergruppe des zweiten Messzeitpunktes. Diese erzielte im Mittel 233,44 Punkte. Da noch keinerlei Befunde vorliegen, welche Auskunft über den Zusammenhang zwischen Solowert und Beurteilung der Verhaltensbeobachtung geben, kann die Annahme, dass die Bewerber und nicht die Bewertungen schlechter geworden sind, zum jetzigen Zeitpunkt nur gemutmaßt werden.

5.4 Beantwortung der Forschungsfragen

Nun gilt es, mit Hilfe der empirischen Befunde der statistischen Analyse die in Kapitel 3 aufgestellten Forschungshypothesen zu beantworten. Die erste Forschungsfrage beschäftigt sich mit der Verwendbarkeit der Ratingdaten. Die entsprechende Hypothese lautet:

Hypothese 1.1 Das DCT-Verfahren erfüllt das Gütekriterium Objektivität. Die Koeffizienten der Interrater-Reliabilität liefern zufriedenstellende Werte.

Die Betrachtung der ICC-Resultate dient der Prüfung dieser Annahme. Aufgrund nichtgegebener Voraussetzung konnte die Hypothese für die Skala *Teamfähigkeit* nicht geprüft werden. Für die restlichen Skalen ergaben sich folgende Resultate.

Die Hypothese 1.1 kann unter Auslassung der Skala *Frustrationstoleranz* angenommen werden. Eine große Anzahl an Beurteilern erhöht die Interrater-Reliabilität des Mittelwertes. Die Average Measure kann somit nicht als Gütemaß für Einzelrater oder Skalen herangezogen werden. Die Beantwortung der Fragestellung orientiert sich demnach an den Ergebnissen der Single Measure. Der Wert liegt hier unterhalb des Grenzwertes 0,4. Somit ist die Beobachterreliabilität bezüglich der *Frustrationstoleranz* nicht akzeptabel. Die Werte der anderen Skalen liefern Resultate mit akzeptabler Ausprägung. In Bezug auf die Verwendbarkeit der Daten ist somit festzustellen, dass die videobasierte Operationalisierung im Kontext des

DCT mit Ausnahme der Skalen *Teamfähigkeit* und *Frustrationstoleranz* psychometrisch verwendbare Daten liefert. Die entsprechenden Koeffizienten der Interrater-Reliabilität ergeben zufriedenstellende Werte, weshalb das Gütekriterium der Objektivität für die Mehrzahl der Skalen erfüllt ist.

Die zweite Forschungsfrage dieser Studie behandelt die Evaluation des Beobachtertrainings. Es soll geklärt werden, ob das Beobachtertraining auf Grundlage der videobasierten Operationalisierung zu einer Verbesserung der Verwendbarkeit der psychometrischen Daten beiträgt. Die erste Hypothese postuliert:

Hypothese 2.1 Das Training führt zu einer Erhöhung der Beobachterübereinstimmung. Die Koeffizienten der Interrater-Reliabilität erhöhen sich.

Die nach Messzeitpunkt differenzierte Betrachtung der ICC als Maß für die Interrater-Reliabilität führt zu folgenden Resultaten. Mit Ausnahme der Skalen *Teamfähigkeit*, *Physiologische Belastbarkeit* und *Kommunikation* resultieren zum zweiten Messzeitpunkt stets höhere Werte. Die Interrater-Reliabilität konnte durch die Beobachterschulung für sechs der neun Ratingskalen erhöht werden. Die größte Verbesserung weisen dabei die Skalen *Impulsivität* und *Dominanz* auf. Die kleineren Vertrauensintervalle nach der Beobachterschulung sind ein weiteres Indiz für eine Erhöhung der Interrater-Reliabilität.

Der zweiten Forschungsfrage untersteht noch eine weitere Hypothese. Diese beinhaltet folgende Annahme:

Hypothese 2.2 Das Training führt zu einer Reduktion von Beobachterfehlern. Zentrale Tendenzen, Milde- bzw. Strengeeffekte werden geringer.

Die deskriptive Betrachtung konnte zeigen, dass vor dem Beobachtertraining die Randnoten 1 und 7 deutlich weniger und zum Teil gar nicht vergeben wurden. Dies spricht für einen Trend zu mittleren Urteilen, bei denen die Skalenenden gemieden werden. Diese Tendenz zur Mitte zeigt sich zum zweiten Messzeitpunkt nach dem Beobachtertraining nicht mehr. Es kann also davon ausgegangen werden, dass das Beobachtertraining zu einer Reduzierung des Beobachterfehlers Tendenz zur Mitte

geführt hat. Tendenzen zu besonders milden beziehungsweise strengen Urteilen konnten weder vor noch nach der Beobachterschulung beobachtet werden. Die Mittelwertsdifferenzen sind stets kleiner als eine Skaleneinheit.

6. DISKUSSION UND AUSBLICK

Im sechsten Abschnitt dieser Diplomarbeit werden die Ergebnisse zunächst einzeln bezüglich der Forschungshypothesen dargelegt. Anschließend erfolgen eine integrative Diskussion und die Erörterung möglicher Ursachen für die entsprechenden Ergebnisse. Der Fokus liegt dabei nicht mehr speziell auf den Hypothesen, sondern setzt an der Betrachtung des großen Ganzen an. Mögliche Einflussfaktoren werden extrahiert und die Ergebnisse im Kontext des Auswahlverfahrens der FFD interpretiert. Im Teil Ausblick werden die Perspektiven des DCT Verfahrens zur Auswahl der FFD aufgezeigt.

6.1 Diskussion der Hypothesen

Die Ergebnisse aus Kapitel 5 werden nun in Hinblick auf die Hypothesen der beiden Forschungsfragen diskutiert. Dabei sollen sowohl Methodik als auch Auswertung dieser Untersuchung einer kritischen Betrachtung unterzogen werden. Mögliche Ursachenzuschreibungen und Alternativerklärungen werden dargelegt.

6.1.1 Forschungsfrage 1

Die erste Forschungsfrage beschäftigt sich mit der Brauchbarkeit der resultierenden Daten. Die vorliegenden Daten stützen, mit Ausnahme der Skalen *Teamfähigkeit* und *Frustrationstoleranz*, die Hypothese 1.1, welche postuliert dass der DCT das Gütekriterium Objektivität erfüllt. Hierfür wurden die $ICC_{unjustiert}$ als Maß für die Interrater-Reliabilität berechnet.

Die gewählte Methodik ließ es nicht zu, dass alle Bewerber von allen Beobachtern bewertet wurden. Somit konnten andere gängigen Übereinstimmungsmaße, wie Cohens κ nicht berechnet werden. Zur Absicherung der Ergebnisse empfiehlt es sich

jedoch, mehrere Koeffizienten zu präsentieren. Ein weiterer Kritikpunkt liegt darin begründet, dass keine allgemeingültigen Kriterien für die Güte einer Beobachterreliabilität vorliegen. Das Vorgehen von Caspar und Wirtz (2002) bei der Überschreitung des Grenzwertes $r=0,4$ von einer akzeptablen und ab $r=0,7$ von einer guten Reliabilität zu sprechen, wird in der Literatur nicht einstimmig empfohlen. Andere Interpretationsstandards setzen die Grenze für eine mittlere Reliabilität bei $r=0,8$. Eine hoch ausgeprägte Reliabilität wäre nach dieser Einteilung erst ab einem Wert von $r=0,9$ interpretierbar (Rindermann, 2002).

Die $ICC_{\text{unjustiert}}$ entspricht der Berechnung einer einfaktoriellen ANOVA. Diese führt bei fehlenden Voraussetzungen und einer kleinen Raterstichprobe zu fehlerhaften Effektschätzungen. Da die Normalverteilung nicht bei allen Skalen vorliegt und die Anzahl der Rater kleiner 10 ist, müssen die Ergebnisse der Interrater-Reliabilität relativiert werden. Parallel dazu bedarf es aus den folgenden Gründen ebenfalls einer Aufwertung der Resultate. Wäre die Berechnung der ICC_{justiert} möglich gewesen, hätte diese zu etwas höheren Werten geführt. Auch die Tatsache, dass die Bewerbergruppe einer relativ homogenen Stichprobe entspricht führt zu einer Unterschätzung der Reliabilität. Dies liegt darin begründet, dass die Reliabilität umso höher ausfällt, je stärker die Mittelwertsunterschiede der Bewerber sind (Caspar & Wirtz, 2002).

Teamfähigkeit erfüllt die Anforderung der Varianzhomogenität nicht, weshalb die $ICC_{\text{unjustiert}}$ in diesem Fall nicht berechnet werden darf. Die Heterogenität könnte durch eine zu kleine Stichprobenwahl begründet werden. Eine detaillierte Analyse der Skala *Teamfähigkeit* steht somit noch aus und muss in Zukunft, anhand weiterer Daten ergänzt werden. So lange die Güte der Skala nicht geklärt werden konnte, dürfen die resultierenden Werte nicht als Bewertungsgrundlage fungieren. Die Skala *Frustrationstoleranz* erzielt als einzige nicht akzeptable Werte bezüglich der Interrater-Reliabilität. Aus diesem Grund sollte die Konstruktion dieser Skala überarbeitet werden. Eventuell sind die Verhaltensanker und Zuweisungsregeln nicht eindeutig genug und müssen weiter spezifiziert werden. Bei den Indikatoren der *Frustrationstoleranz* handelt sich um sehr vielschichtige Verhaltensweisen, die sich schwer aus den Bearbeitungsverlauf lösen lassen. Vielleicht übersteigt die Komplexität der Bewertung, in Summe mit derer der anderen Skalen, die Konzentrationsspanne der Rater und sollte vereinfacht oder gar vernachlässigt

werden. Auch hier gilt, in der jetzigen Form dürfen die Resultate der Ratingskala keinesfalls in die Bewertung einfließen oder interpretiert werden.

6.1.2 Forschungsfrage 2

Die zweite Forschungsfrage setzt sich mit der Evaluation des Beobachtertrainings auseinander. Die erste Hypothese postuliert, dass die Beobachterübereinstimmung durch das Training erhöht wird und somit die Koeffizienten der Interrater-Reliabilität steigen. Die Hypothese 2.1 konnte lediglich für einige Skalen angenommen werden, da sich die Beobachterreliabilität zum Messzeitpunkt zwei ausschließlich für die Skalen *subjektive Belastbarkeit*, *Impulsivität vs. Reflexivität*, *Frustrationstoleranz*, *Flexibilität*, *Dominanz* und *kognitive Leistung* erhöhte. Dabei war der Sprung bei *Dominanz* und *Impulsivität vs. Reflexivität* besonders hoch. Dies könnte mit der Struktur der Ratingskalen im Zusammenhang stehen. Es handelt sich hier um die eher unüblichen Doppelskalen, deren Handhabung auf den ersten Blick etwas ungewohnt scheint. Im Gegensatz dazu ist der Zugang zu den einfachen Likert-Skalen, gerade für Personen mit psychologischer Grundausbildung, wohl intuitiver. So kann die starke Verbesserung in diesen beiden Bereichen auch dadurch erklärt werden, dass die Konstruktion der Skalen zunächst weniger verstanden wurde und die Anwendung an sich erst eingeübt werden musste.

Der DCT basiert auf einer wissenschaftlichen Verhaltensbeobachtung und stellt sehr große Anforderungen an die Rater. Im Kontext einer komplexen Problemlöseaufgabe werden neun Skalen gleichzeitig für zwei verschiedene Bewerber bewertet. Dies fordert ein extrem hohes Maß an Konzentration, die über sechs Teamrunden hinweg aufrechterhalten werden muss. Da es sich hierbei um eine standardisierte Situation handelt, in der bestimmte Verhaltensweisen gehäuft gezeigt werden, tritt mit der Zeit eine gewisse Beobachtungsroutine ein. Die Rater werden immer vertrauter mit der Aufgabe und können sich mehr auf die einzelnen Verhaltensweisen konzentrieren. In der anschließenden Beobachterkonferenz werden Grenzfälle besprochen und Meinungsverschiedenheiten diskutiert. Diese Lern- und Annäherungseffekte könnten eine zusätzliche Erklärung für die Verbesserung der Interrater-Reliabilität sein. Für die Skalen *physiologische Belastbarkeit* und *Kommunikation* sank das Ausmaß der Beobachterreliabilität. Vermutlich hat das Training in diesen beiden Fällen Uneinigkeit unter den Ratern geschaffen. Im Kontext der *physiologischen Belastbarkeit* könnte dies mit einer unscharfen Abgrenzung der

handlungsbegleitenden und handlungssteuernden Indikatoren in Verbindung stehen. Diesbezügliche Definitionen und Indikatoren sollten spezifiziert werden. Ein anderer Erklärungsansatz begründet sich durch die bereits vorhandenen Auswahlverfahren des Dezernats. Die Belastbarkeit der Bewerber wird auch an anderen Stellen erhoben und bewertet. Dies geschieht jedoch losgelöst vom DCT auf Grundlage andersartiger Verhaltensanker. Solche Überlappungen könnten zu Unsicherheiten geführt haben. Ist man vor dem Training noch einig über Klassifikation und Quantifizierung der in anderen Bereichen üblichen Belastungsanzeichen, führt die Schulung der spezifischen DCT-Indikatoren möglicherweise zu Verwirrungen und Uneinigkeit. Ähnlich könnte es sich auch bezüglich der Skala *Kommunikation* verhalten. Dieser allgemein gebräuchliche Begriff beschreibt im DCT-Szenario ein sehr umgrenztes Verhaltensspektrum. Vielleicht hat auch hier das Training eher zu Verwirrungen als zur erwünschten Einigkeit geführt. Um die entstandenen Unsicherheiten zu lösen, ist eine Nachschulung dieser beiden Skalen dringend von Nöten. Auch eine gemeinsame Diskussion könnte Missverständnisse aufzeigen und eliminieren.

Eine weitere Hypothese der Forschungsfrage zwei nimmt an, dass die Beobachterfehler Tendenz zur Mitte, Milde- und Strengeeffekte durch das Beobachtertraining reduziert wurden. Die Daten zeigten zu keinem Zeitpunkt Milde- oder Strengeeffekte der Beurteiler. Die Tendenz zu Mitte zeichnete sich hingegen zum ersten Messzeitpunkt ab und konnte nach dem Beobachtertraining verringert werden. Boden und Deckeneffekte, wie sie sich in den Daten des ersten Messzeitpunkts abzeichnen, treten im Zusammenhang mit Ratingskalen häufig auf. Unsicherheit in der Bewertung begünstigt diese zentralen Tendenzen. Ein weiterer Indikator für Unsicherheit in der Beurteilung ist der häufige Gebrauch der Antwortalternative „nicht beobachtet“. Beide Indizien für Unsicherheit in der Bewertung konnten nach dem Durchlaufen des Beobachtertrainings reduziert werden. Jedoch könnte auch hier die gegenseitige Absprache innerhalb der Kommission einen Einfluss haben.

6.2 Integrative Diskussion

Im vorherigen Teil wurden die Ergebnisse auf Ebene der einzelnen Hypothesen diskutiert. Nun folgt eine Gesamtbetrachtung losgelöst von dieser Einteilung. Insgesamt kann festgestellt werden, dass die Konstruktion der Skalen in den meisten

Fällen gut gelungen ist. Viele der Skalen konnten durch das Beobachtertraining zusätzlich optimiert werden. Es ist also möglich mit Hilfe des DCTs psychometrisch verwendbare und objektive Daten bezüglich der Skalen *physiologische Belastbarkeit*, *subjektive Belastbarkeit*, *Impulsivität*, *Flexibilität*, *Kommunikation*, *Dominanz* und *kognitive Leistung* zu erheben. Diese Dimensionen können die in der Einleitung aufgezeigten Lücken schließen und einen erheblichen Beitrag zum diagnostischen Mehrwert des gesamten Auswahlverfahrens liefern. Dabei kann der DCT keines der jetzigen Verfahren ersetzen, denn gerade in der Personalauswahl ist ein multimodulares Vorgehen für einen umfassenden Erkenntnisgewinn unabdingbar (Kleinmann, 2003).

Da der DCT zur Bewerberauswahl im Kontext der FFD eingesetzt werden soll, ist es ratsam die Ergebnisse dieser Studie mit anderen Methoden zusätzlich abzusichern. Zur genaueren Prüfung der Reliabilitätskoeffizienten sollten mehrere Rater denselben Fall bewerten. Dies könnte zum Beispiel über die Beurteilung von Beispielfideos gelöst werden. Somit ist es möglich spezielle Defizite bestimmter Skalen zu extrahieren und die Rater in diesen Bereichen einer gezielten Nachschulung zu unterziehen. Durch die Einbeziehung der Bewerbermimik sind Steigerungen des Informationsgewinns und somit des diagnostischen Mehrwerts durchaus denkbar. Das Anbringen eines Wandspiegels könnte eine zusätzliche Perspektive auf den Gesichtsausdruck der Bewerber schaffen. Dennoch muss an dieser Stelle eingeräumt werden, dass die menschliche Wahrnehmungsfähigkeit begrenzt ist. Wird die Kapazität überschritten, sind die Daten einer Verhaltensbeobachtung wertlos. Aus diesem Grund wäre es zudem sinnvoll jedem Beobachter die Verantwortung für einen Bewerber zu übertragen. So kann die Komplexität der Verhaltensbeobachtung verringert werden.

6.3 Ausblick

Diese Arbeit verdeutlicht, dass der DCT, mit Ausnahme zweier Skalen, psychometrisch verwendbare Daten liefert und das Gütekriterium der Objektivität erfüllt. Das konstruierte Beobachtertraining optimiert diese Eigenschaften. Folglich besitzt das Verfahren großes Potential für die Auswahl der FFD am Dezernat Auswahl der Fachgruppe Flugpsychologie am Zentrum für Luft- und

Raumfahrtmedizin der Luftwaffe. Durch den persönlichen Kontakt zwischen Bewerber und Psychologe weist das Verfahren ein großes Ausmaß an Transparenz und Augenscheinvalidität auf. Dies bietet die Möglichkeit eines informativen Feedbacks und einer hohen Akzeptanz seitens der Bewerber (Kleinmann, 2013). Zusätzlich ermöglicht es einen intensiven Informationsgewinn, welcher den Auswahlprozess weiter optimieren, und einen maßgeblich zur Entscheidungsfindung beitragen kann.

Nach den Betrachtungen dieser Studie bleiben noch einige Fragen offen. Objektivität bildet ein zentrales Gütekriterium der psychologischen Diagnostik. Neben der bisher betrachteten Auswertungsobjektivität ist die objektive Verfahrensdurchführung ebenfalls ein fundamentales Kriterium der Güte eines Auswahlverfahrens. Für vergleichbare Resultate sollte die Durchführung weitestgehend standardisiert ablaufen. Es stellt sich die Frage, inwieweit jede Person gerecht beurteilt werden kann, wenn die Beobachtung abhängig von der Interaktion zweier zueinander in Konkurrenz stehender Personen ist. Da im DCT zwei Bewerber interagieren, ist ein standardisierter Ablauf nicht in allen Fällen garantiert. Durch die wechselnden Eigenschaften und Verhaltensweisen des zufällig zugeteilten Partners kann das Handeln der Bewerber nur begrenzt verglichen werden. Ein Gedankenexperiment verdeutlicht, dass ein Bewerber mit durchschnittlichen kommunikativen Fähigkeiten in Zusammenarbeit mit einem sehr guten Teilnehmer womöglich bessere Ergebnisse erzielen wird, als mit einer Person die unterdurchschnittliche Kommunikationsfähigkeiten besitzt. Gerade bei Partnern in extremen Leistungsbereichen birgt dies die Gefahr starker Verzerrungen. Eine Lösungsmöglichkeit bildet der Einsatz von bundeswehrinternen Bearbeitungspartnern, welche anhand genauer Vorgaben stets in der gleichen Art und Weise agieren und reagieren. So kann die Testsituation zusätzlich standardisiert werden. Dem gegenüber steht jedoch das Gütekriterium der Ökonomie. Folgt man wirtschaftlichen Überlegungen sollte, gemessen am resultierendem Informationsgewinn, möglichst wenig Ressourcen in Anspruch genommen werden. Zeit und Geld bilden in Hinblick auf den Auswahlprozess der Bundeswehr die zentralen Ressourcen und sind nur begrenzt verfügbar. Der DCT ist mit einer Gesamtdurchführungszeit von über einer Stunde ein sehr zeitintensives Verfahren. Die Einzelbetrachtung eines jeden Bewerbers und die Anstellung standardisierten Bearbeitungspartnern könnte nur eine entsprechende Steigerung des Informationsgewinns rechtfertigen. Zur Optimierung des Verfahrens sollte dieser

Fragestellung weiter nachgegangen und die geeignete Balance zwischen Ökonomie und Aufrechterhaltung der Sachlichkeit ermittelt werden. Bis dies geschehen ist, muss darauf geachtet werden die Bewerber möglichst losgelöst zu betrachten und Bewerber mit ähnlichem Eignungsgrad zu paaren.

Aus testökonomischen Gründen wäre eine Zusammenfassung durchaus sinnvoll. Anschließende Analysen sollen zeigen, ob und wie die Skalen gewichtet und somit zu einem Gesamtwert verrechnet werden können. Des Weiteren ist noch offen, welche Bedeutung dem Solowert der Einzelaufgaben zukommt. Kommt man zu dem Entschluss den Solowert in die Bewertung des DCTs zu integrieren, erfordert dies eine ausführliche Normierung. Des Weiteren stellt sich die Frage, wie die Skalen untereinander und mit den Ergebnissen der Vorrunde korrelieren. Die Resultate könnten Hinweise darauf geben, wie das Abschneiden im DCT und kognitiven Fähigkeiten einer Person in Verbindung stehen. Da das Verfahren sehr zeitaufwändig ist, würden derartige Erkenntnisse den organisatorischen Prozess erleichtern. Durch die vorangestellte Testung der kognitiven Eignung könnten ungeeignete Bewerber schon vor der Bearbeitung des DCT abgelöst werden. Dies würde die Verfahrenseffektivität steigern.

Ein wissenschaftlich fundierter Einsatz erfordert einige Folgestudien. Gütesicherung und Evaluation sollten ständige Begleiter eines jeden Verfahrens sein. Die prognostische Validität kann erst nach fortgeschrittener Ausbildung der Bewerber erhoben werden. Entsprechende Analysen folgen zu einem späteren Zeitpunkt mit den Daten aus der Langzeiterfolgskontrolle (LEK). Die Konstruktvalidität bildet einen weiteren Anteil der Gesamtvalidität. Sie kann mit Hilfe von Expertenratings belegt werden. Zusätzlich sollten die Verhaltensanker der Indikatorliste ständig ergänzt und von der Kommission diskutiert werden.

Der DCT muss als dynamisches Segment im multimodularen Auswahlverfahren angesehen werden, dessen Konstruktion und Gütesicherung ein ständig fortlaufender Prozess ist.

7. ZUSAMMENFASSUNG

Anforderungsanalysen und Flugunfalluntersuchungen zeigen die Notwendigkeit nicht kognitiver Fähigkeiten in den Tätigkeitsbereichen der Flugsicherung. Im Kontext theoretischer Betrachtungen werden neben den kognitiven Fähigkeiten einer Person häufig die Bereiche Belastbarkeit, Arbeitsstil und Interaktion als fundamentale Leistungsdeterminanten angeführt. Der Dyadic Cooperation Test bietet die Möglichkeit diese Dimensionen in den Auswahlprozess der Flugführungsdienste des Dezernats Auswahl der Fachgruppe Flugpsychologie am Zentrum für Luft- und Raumfahrtmedizin der Luftwaffe zu integrieren. Für den Einsatz des Verfahrens mussten die theoretischen Konstrukte zunächst operationalisiert und im Anschluss bezüglich ihrer Güte geprüft werden. Diese Methodik bildet den Inhalt der vorliegenden Diplomarbeit. Mit dem Anspruch möglichst hoher Transparenz werden die einzelnen Konstruktionsschritte: videobasierte Generierung der Indikatoren, Skalenklassifikation und Quantifizierung sowie die Prüfung der Verfahrensgüte im Detail beschrieben. Außerdem wird die Konstruktion einer Beobachterschulung dargelegt. Anhand der statistischen Auswertung von 63 Bewerberratings kann gezeigt werden, dass die konstruierten Skalen *physiologische Belastbarkeit*, *subjektive Belastbarkeit*, *Impulsivität*, *Flexibilität*, *Kommunikation*, *Dominanz* und *kognitive Leistung* objektive und somit psychometrisch verwendbare Daten liefern. Bezüglich der Skalen *Teamfähigkeit* und *Frustrationstoleranz* bedarf ein wissenschaftlich fundierter Einsatz noch an Feinabstimmung. Die Schulung der Rater mit Hilfe des erstellten Beobachtertrainings führt zudem in den meisten Fällen zur Verbesserung der Objektivität. Obwohl Folgestudien wie die Validitätsprüfungen noch ausstehen, kann mit dieser Diplomarbeit gezeigt werden, dass die Überarbeitung des DCT das Potential besitzt in den Auswahlprozess eingegliedert zu werden und diesen somit zu optimieren.

Physische Belastungsanzeichen

Video	Absatz Verhaltensanker	Subkategorie	Ausprägung
TRANS_DCT_Runde_1	41 angespannte/verkrampfte Sitzhaltung	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_1	215 knackt Finger	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_2	79 Hände verkrampft	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_3	46 gebeugte Körperhaltung	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_5	85 Finger verschränkt	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_5	105 schüttelt Hand aus	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_5	240 zittert	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_7	69 stottert	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_6	63 sitzt relativ starr	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_6	81 faltet Hände auf Tisch	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_6	82 räuspert sich	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_7	170 stützt Hände an die Taille, streckt Rücken	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_8	123 schlägt Hände vor dem Gesicht zusammen	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_9	30 pustet Backen auf	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_9	126 Faust vor Mund	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_10	93 ballt Fäuste	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_2	24 atmet laut aus	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_2	128 schnauft	Anspannung	handlungsbegleitend
TRANS_DCT_Runde_3	280 Gesicht gerötet	Durchblutung	handlungsbegleitend
TRANS_DCT_Runde_2	8 sitzt ruhig	keine Belastung	handlungsbegleitend
TRANS_DCT_Runde_4	111 ruhige Stimme	keine Belastung	handlungsbegleitend
TRANS_DCT_Runde_8	4 Hände liegen ruhig auf dem Tisch	keine Belastung	handlungsbegleitend
TRANS_DCT_Runde_10	6 sitzt aufrecht	keine Belastung	handlungsbegleitend
TRANS_DCT_Runde_10	13 klare und hohe Stimme	keine Belastung	handlungsbegleitend
TRANS_DCT_Runde_1	23 dreht sich auf Stuhl hin und her	keine Belastung	handlungsbegleitend
TRANS_DCT_Runde_1	35 Beine unruhig	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_1	61 greift sich ins Gesicht	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_2	4 Hand am Mund	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_2	10 Hand an Kinn	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_2	98 Greift sich an den Kopf	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_3	45 rutscht auf Stuhl hin und her	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_3	124 klopf mit Fingern auf Tisch	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_5	63 zupft an Kleidung	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_5	218 kaut auf Fingernagel	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_1	204 tippt mit Fingern auf Tisch	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_6	64 nimmt Hand zum Oberkörper	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_7	55 wackelt mit Finger	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_7	130 spielt mit Stift	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_7	211 fasst sich an den Hals	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_7	248 fasst sich an Schulter	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_8	36 schnipst mit Finger	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_8	74 rutscht mit Stuhl zurück	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_8	75 wischt Hände an Pullover	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_9	34 beißt auf Lippe	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_10	71 Hand an Kopfhörerkabel	motor. Unruhe	handlungsbegleitend
TRANS_DCT_Runde_10	109 kratzt sich am Kopf	Anspannung	handlungssteuernd
Theoretische Überlegung	Ohnmacht	Anspannung	handlungssteuernd
Theoretische Überlegung	Schockschreie	Anspannung	handlungssteuernd
Theoretische Überlegung	fällt vom Stuhl	motor. Unruhe	handlungssteuernd
Theoretische Überlegung	kann nicht reden	Anspannung	handlungssteuernd

ANHANG

Indikatorlisten

Subjektive Belastungsanzeichen

Video	Absatz	Verhaltensanker	Subkategorie	Ausprägung
TRANSS_DCT_Runde_1	158	äußert Unsicherheit	Emotion	handlungsbegleitend
TRANSS_DCT_Runde_1	227	lächelt	Emotion	handlungsbegleitend
TRANSS_DCT_Runde_3	61	kichert	Emotion	handlungsbegleitend
TRANSS_DCT_Runde_5	23	schaut kurz zu Beobachtern	Emotion	handlungsbegleitend
TRANSS_DCT_Runde_1	100	rechnet laut	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_1	163	redet mit sich selbst	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_3	44	verhaspelt sich	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_3	134	murmelt Zahlen	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_7	118	rechnet laut	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_8	103	nutzt Finger zum Merken	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_8	141	murmelt zahlen, bewegt Hände rhythmisch mit	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_8	166	zählt seine Kapazitäten auf, nutzt dafür seine Finger,	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_10	140	erfragt „Gewichtszustände“ (richtige Wort fällt ihr nicht ein	Konzentration	handlungsbegleitend
TRANSS_DCT_Runde_6	15	beide Hände am Nummernblock	Koordination	handlungsbegleitend
TRANSS_DCT_Runde_6	68	Eingabe (benutzt beide Hände)	Koordination	handlungsbegleitend
TRANSS_DCT_Runde_3	21	drückt Tasten sehr fest	Koordination	handlungsbegleitend
TRANSS_DCT_Runde_6	93	zieht Nummernblock sehr nah zu sich	Koordination	handlungsbegleitend
TRANSS_DCT_Runde_1	90	zeigt mit Finger auf Bildschirm	Wahrnehmung	handlungsbegleitend
TRANSS_DCT_Runde_1	104	Finger am Bildschirm	Wahrnehmung	handlungsbegleitend
TRANSS_DCT_Runde_3	27	beugt sich weit nach vorne	Wahrnehmung	handlungsbegleitend
TRANSS_DCT_Runde_5	7	bleibt mit Finger am Bildschirm, während der Eingabe	Wahrnehmung	handlungsbegleitend
TRANSS_DCT_Runde_7	35	bewegt Kopf beim lesen mit	Wahrnehmung	handlungsbegleitend
Theoretische Überlegung		Denkblockaden	Konzentration	handlungssteuernd
Theoretische Überlegung		Blackouts	Wahrnehmung	handlungssteuernd
Theoretische Überlegung		Keine Eingabe möglich	Koordination	handlungssteuernd
Theoretische Überlegung		Lachkrampf	Emotion	handlungssteuernd

Impulsivität

Video	Absatz	Verhaltensanker	Subkategorie	Ausprägung
TRANS_DCT_Runde_3	77	überlegt während des Sprechens	vorschnell	impulsiv
TRANS_DCT_Runde_3	132	beginnt Kommunikation, bricht sie dann aber wieder ab	vorschnell	impulsiv
TRANS_DCT_Runde_3	188	wartet Antwort nicht ab und gibt sofort neue Meldung	vorschnell	impulsiv
TRANS_DCT_Runde_3	246	antwortet, obwohl Zeit abgelaufen	vorschnell	impulsiv
TRANS_DCT_Runde_7	45	beginnt Eingabe, unterbricht für Sprechen	vorschnell	impulsiv
TRANS_DCT_Runde_7	136	Eingabe schnell	vorschnell	impulsiv
TRANS_DCT_Runde_7	171	rechnet während des Sprechens	vorschnell	impulsiv
TRANS_DCT_Runde_8	105	erste Meld. Sehr schnell	vorschnell	impulsiv
TRANS_DCT_Runde_4	21	erste Meldung erst nach kurzer Wartezeit	zögerlich	reflexiv
TRANS_DCT_Runde_4	25	gibt Zahlen ein, wartet etwas bis er bestätigt	zögerlich	reflexiv
TRANS_DCT_Runde_5	20	arbeiten sehr langsam	zögerlich	reflexiv
TRANS_DCT_Runde_5	35	bricht Eingabe ab, entscheidungsscheu	zögerlich	reflexiv
TRANS_DCT_Runde_6	10	beide warten lange ab	zögerlich	reflexiv
TRANS_DCT_Runde_6	47	Bearbeitung sehr bedacht und zögerlich	zögerlich	reflexiv
TRANS_DCT_Runde_6	58	tut sehr lange nichts	zögerlich	reflexiv
TRANS_DCT_Runde_6	72	wartet, beginnt Eingabe, zögert, löscht sie wieder, beginnt neue Eingabe	zögerlich	reflexiv
TRANS_DCT_Runde_7	41	zögert, kreist mit Fingern über Nummernblock	zögerlich	reflexiv
TRANS_DCT_Runde_8	62	Eingabe sehr langsam	zögerlich	reflexiv
TRANS_DCT_Runde_10	56	Tatenlos, schaut auf Bildschirm	zögerlich	reflexiv
TRANS_DCT_Runde_1	63	"versuche irgendwie zu verteilen, bis Fehlermeldung kommt"	probieren	impulsiv
TRANS_DCT_Runde_2	33	erneute Eingabe, wieder Überlastungsanzeige	probieren	impulsiv
TRANS_DCT_Runde_3	55	Anweisung, Ost soll es mit 6 E versuchen	probieren	impulsiv
TRANS_DCT_Runde_9	112	„Ich glaub das haut so ein bisschen hin“ --> Ende	probieren	impulsiv
TRANS_DCT_Runde_5	74	Antwortet ohne Sprechentaste	Flüchtigkeitsfehler	impulsiv
TRANS_DCT_Runde_5	23	drückt ungewollt Endetaste	Flüchtigkeitsfehler	impulsiv
TRANS_DCT_Runde_6	6	drückt Sprechentaste doppelt - geht wieder aus	Flüchtigkeitsfehler	impulsiv
TRANS_DCT_Runde_10	94	drückt Anfangstaste ohne Absprache	Flüchtigkeitsfehler	impulsiv
TRANS_DCT_Runde_7	140	vertippt sich	Flüchtigkeitsfehler	impulsiv
TRANS_DCT_Runde_10	39	keine Korrekturen	Korrektur	reflexiv
TRANS_DCT_Runde_1	159	ergänzt Problemmeldung nach 1. Meld	Korrektur (mündlich)	impulsiv
TRANS_DCT_Runde_2	18	kündig Korrektur auf Str. 1 an	Korrektur (mündlich)	impulsiv
TRANS_DCT_Runde_2	42	ergänzt fehlende erste Meldung	Korrektur (mündlich)	impulsiv
TRANS_DCT_Runde_4	29	korrigiert Verhältnisangabe seiner Str.	Korrektur (mündlich)	impulsiv
TRANS_DCT_Runde_7	65	entschuldigt sich, habe sich versprochen	Korrektur (mündlich)	impulsiv
TRANS_DCT_Runde_8	183	vergisst 1. Meld., korrigiert sich	Korrektur (mündlich)	impulsiv
TRANS_DCT_Runde_7	242	beginnt Eingabe, korrigiert sie dann	Korrektur (Eingabe)	impulsiv
TRANS_DCT_Runde_3	56	drückt Korrekturtaste	Korrektur (Eingabe)	impulsiv

Regelkonformität

Video	Absatz	Verhaltensanker	Subkategorie	Ausprägung
TRANS_DCT_Runde_1	139	1. Meldung nicht korrekt (Probleme fehlen)	nicht regelkonform	impulsiv
TRANS_DCT_Runde_2	116	1. Meldung nicht korrekt (sagt nicht "Hier spricht City ...")	nicht regelkonform	impulsiv
TRANS_DCT_Runde_2	130	1. Meldung nicht korrekt (Verantwortung fehlt)	nicht regelkonform	impulsiv
TRANS_DCT_Runde_6	127	1. Meld. nicht korrekt (falsche Formulierung, „ich bin City Ost“)	nicht regelkonform	impulsiv
TRANS_DCT_Runde_8	183	vergisst 1. Meld.	nicht regelkonform	impulsiv
TRANS_DCT_Runde_1	7	1. Meldung, korrekt	regelkonform	effektiv

Frustration			
Video	Absatz	Verhaltensanker	Subkategorie
TRANS_DCT_Runde_1	27	Zeit abgelaufen stöhnt	Frustration
TRANS_DCT_Runde_1	99	Überlastung, „häh“ und erstaunte Gestik	Frustration
TRANS_DCT_Runde_1	115	Zeit läuft ab, schüttelt mit Kopf	Frustration
TRANS_DCT_Runde_1	138	Zeit abgelaufen: „Mist“	Frustration
TRANS_DCT_Runde_1	151	Überlastung: seufzt	Frustration
TRANS_DCT_Runde_1	196	Zeit läuft ab, schnaut	Frustration
TRANS_DCT_Runde_2	11	Zeit abgelaufen, geht mit Körper etwas zurück	Frustration
TRANS_DCT_Runde_2	32	Überlastungsanzeige, flucht	Frustration
TRANS_DCT_Runde_2	38	Zeit abgelaufen: flucht	Frustration
TRANS_DCT_Runde_2	95	Überlastung: schnaut	Frustration
TRANS_DCT_Runde_2	142	Zeitabgelaufen: klopft Hand leicht auf Tisch	Frustration
TRANS_DCT_Runde_4	54	Zeit abgelaufen: Räuspert sich und weicht zurück	Frustration
TRANS_DCT_Runde_4	138	Zeit abgelaufen, Hand vorm Mund, schluckt	Frustration
TRANS_DCT_Runde_5	10	Zeit abgelaufen, Hände verschränkt, nach vorne gebeugt	Frustration
TRANS_DCT_Runde_5	40	Zeit abgelaufen, nimmt Hände in die Luft und lässt sie auf Tisch fallen	Frustration
TRANS_DCT_Runde_6	91	Zeit abgelaufen, nimmt Faust vor den Mund	Frustration
TRANS_DCT_Runde_7	53	Überlastung: murmelt verwundert	Frustration
TRANS_DCT_Runde_7	96	schauf bei Überlastung schnell auf Bildschirm	Frustration
TRANS_DCT_Runde_9	46	Zeit abgelaufen „toll“	Frustration
TRANS_DCT_Runde_9	79	Zeit abgelaufen: lässt Kopf hängen	Frustration
TRANS_DCT_Runde_9	130	kann Belastungsanzeige nicht sehen „oh Gott“,	Frustration
TRANS_DCT_Runde_10	105	Zeit abgelaufen: greift sich an den Kopf	Frustration

Flexibilität

Video	Absatz	Verhaltensanker	Subkategorie	Ausprägung
Theoretische Überlegung		Koordinator: gibt Verantwortung ab, leitet Zuarbeiter nicht an	nicht rollenkonform	unflexibel
Theoretische Überlegung		Zuarbeiter: übernimmt Koordination, führt Anweisungen nicht aus	nicht rollenkonform	unflexibel
TRANS_DCT_Runde_1	122	Koordinator: fragt oft nach, nutzt Konjunktiv, ungenaue Formulierungen	nicht rollenkonform	unflexibel
TRANS_DCT_Runde_1	123	Zuarbeiter: gibt viele Vorschläge, legt fest was er als nächstes macht	nicht rollenkonform	unflexibel
TRANS_DCT_Runde_1	166	Zuarbeiter: entscheidet selbst über seinen nächsten Schritt und informiert Ost	nicht rollenkonform	unflexibel
TRANS_DCT_Runde_9	168	Zuarbeiter: sagt was er tun soll (nicht regelkonform)	nicht rollenkonform	unflexibel
Theoretische Überlegung		Koordinator: übernimmt Koordination, gibt Anweisungen,	rollenkonform	flexibel
Theoretische Überlegung		Zuarbeiter: Unterstützt, führt Anweisungen aus,	rollenkonform	flexibel
TRANS_DCT_Runde_8	148	Systemausfall, Anweisung immer Rückmeldung wenn neue besetzt	Systemausfall	flexibel
TRANS_DCT_Runde_7	259	verneint, überlegt warum Fahrzeuge nicht aufgehen, liest nochmal, sagt er überlege kurz	unlösbare Aufgabe	unflexibel

Dominanz

Video	Absatz	Verhaltensanker	Subkategorie
TRANS_DCT_Runde_3	271	wiederholt fragend, 11?	Absicherung
TRANS_DCT_Runde_8	114	fragt ob seine Umrechnung richtig	Absicherung
TRANS_DCT_Runde_8	129	wiederholt 5-10-40-5	unterwürfig
TRANS_DCT_Runde_7	29	Ost hilft erst aus und beginnt erst später mit seinen Straßen	unterwürfig
TRANS_DCT_Runde_3	50	bearbeitet eigenverantwortliche Str. nicht, fragt was Ost auf Str. 1 braucht	unterwürfig
TRANS_DCT_Runde_4	58	tut nichts, wartet auf Anweisung	unterwürfig
TRANS_DCT_Runde_1	26	hat nur noch 20 „Kilo“, braucht sie aber selbst für Str. 1, „soll ich die auf Str. 1 packen oder möchtest du sie verteilen“ - <u>Selbstlos</u>	unterwürfig
TRANS_DCT_Runde_8	21	beginnt Eingabe für seine Str. und beendet diese, während West spricht	unterwürfig
TRANS_DCT_Runde_1	6	meldet sich zuerst	bestimmend
TRANS_DCT_Runde_1	13	Anweisung imperativ	bestimmend
TRANS_DCT_Runde_1	74	zählt ein	bestimmend
TRANS_DCT_Runde_1	102	gibt genaue Anweisung	bestimmend
TRANS_DCT_Runde_1	58	geht nicht auf Frage ein	bestimmend
TRANS_DCT_Runde_2	45	führt nicht aus, gibt erst eigene Anweisung, imperativ	bestimmend
TRANS_DCT_Runde_4	202	zählt Ende ein	bestimmend
TRANS_DCT_Runde_8	24	stellt erst eigene Eingabe fertig	bestimmend
TRANS_DCT_Runde_9	48	drückt Start ohne Einzahlen	bestimmend

Kooperation

Video	Absatz	Verhaltensanker	Subkategorie
TRANS_DCT_Runde_1	57	beide arbeiten für sich	Einzelarbeit
TRANS_DCT_Runde_9	48	drückt Start ohne Einzahlen, Anfangsbildschirm erscheint, nach kurzer Zeit Oh, 3-2-1-Ios Obwohl schon gestartet	Einzelarbeit
TRANS_DCT_Runde_7	165	gibt zur gleichen Zeit schon 5 ein - mangelnde Absprache	Einzelarbeit
TRANS_DCT_Runde_1	130	geht nicht auf Nachfrage ein	Einzelarbeit
TRANS_DCT_Runde_2	27	führt ohne Bestätigung aus	Einzelarbeit
TRANS_DCT_Runde_6	87	keine Antwort, Eingabe	Einzelarbeit
TRANS_DCT_Runde_7	117	beide arbeiten für sich	Einzelarbeit
TRANS_DCT_Runde_9	129	Kümmert sich zunächst alleine um Str. 1	Einzelarbeit
TRANS_DCT_Runde_10	58	beide arbeiten für sich, sprechen nicht miteinander	Einzelarbeit
TRANS_DCT_Runde_1	43	fragt nach Hilfe	erfragt Hilfe
TRANS_DCT_Runde_1	92	macht Vorschläge auf welche Straße er Gewichte verteilen könnte	spezifische Hilfe
TRANS_DCT_Runde_1	180	Überlastung - reagiert sofort und gibt aktuellen Belastungsstand	spezifische Hilfe
TRANS_DCT_Runde_2	153	sagt Belastung unaufgefordert	spezifische Hilfe
TRANS_DCT_Runde_3	48	gibt Alternativvorschlag,	spezifische Hilfe
TRANS_DCT_Runde_3	212	verweist auf die fehlende Str. 4	spezifische Hilfe
TRANS_DCT_Runde_7	228	sagt Belastung und erinnert an Beschränkung Str. 3	spezifische Hilfe
TRANS_DCT_Runde_8	48	schlägt vor Str. 4 anders zu belegen	spezifische Hilfe
TRANS_DCT_Runde_9	122	versucht sie zu unterstützen und fragt welche Straßen schon voll sind	spezifische Hilfe
TRANS_DCT_Runde_10	113	„ich kann Str. 4 noch mit 2 E belegen“, Vorschlag, direkte Hilfe	spezifische Hilfe
TRANS_DCT_Runde_3	72	fragt erneut nach welche Str. (Weist durch Nachfragen auf Fehler hin)	spezifische Hilfe
TRANS_DCT_Runde_2	35	schlägt Kompromiss vor	Strategieabsprache
TRANS_DCT_Runde_3	179	will, dass beide Str. 2 gemeinsam bearbeiten	Strategieabsprache
TRANS_DCT_Runde_3	273	kündigt eigene Handlung an	Strategieabsprache
TRANS_DCT_Runde_4	24	schlägt vor, sich erst auf seine Verantwortlichkeiten zu konzentrieren und dann mit Rest bei Ost auszuhelfen	Strategieabsprache
TRANS_DCT_Runde_5	41	fragt ob West bereit - bjaht - Einzahlen	Strategieabsprache
TRANS_DCT_Runde_6	67	gibt Verantwortung für Str. 1 ab	Strategieabsprache
TRANS_DCT_Runde_8	148	Systemausfall, Anweisung immer Rückmeldung wenn neue besetzt	Strategieabsprache
TRANS_DCT_Runde_1	29	beide reden viel miteinander, Sprechanteile gleich	Team
TRANS_DCT_Runde_1	116	sprechen sich ständig und genau ab, arbeiten kaum für sich	Team
TRANS_DCT_Runde_1	192	versuchen Problem gemeinsam zu lösen	Team
TRANS_DCT_Runde_3	33	bedankt sich	Team
TRANS_DCT_Runde_3	230	sagt, er wartet auf Anweisung	Team
TRANS_DCT_Runde_1	149	widerspricht respektvoll	Team
TRANS_DCT_Runde_7	207	sagt erneut Str. 4 nur zur Hälfte belegbar (weist auf Fehler hin)	Team
TRANS_DCT_Runde_8	189	gibt Feedback - "sehr gut gemacht"	Team
TRANS_DCT_Runde_9	59	Bitet um 4 auf Str. 1	Team
TRANS_DCT_Runde_10	36	spricht West mit Sie an	Team
TRANS_DCT_Runde_1	51	sagt wie viele Typen und Punkte er noch hat	unspezifische Hilfe
TRANS_DCT_Runde_4	35	fragt wo Ost noch Gewichte braucht	unspezifische Hilfe
TRANS_DCT_Runde_4	172	bestätigt, Str. 1 fertig und fragt ob er noch eine Str. bearbeiten soll	unspezifische Hilfe
TRANS_DCT_Runde_7	90	fragt wo sie Ost helfen soll	unspezifische Hilfe

Kommunikation

Video	Absatz	Verhaltensanker	Subkategorie
TRANS DCT Runde 3	119	meint, sie haben sich falsch verstanden	versteht Partner nicht
TRANS DCT Runde 1	11	stottern	unverständlicher Ausdruck
TRANS DCT Runde 1	16	spricht mit Hand vorm Mund	unverständlicher Ausdruck
TRANS DCT Runde 1	20	spricht undeutlich und abgehackt	unverständlicher Ausdruck
TRANS DCT Runde 1	109	fragt: „hast du noch mehr von dem?“ (ungenau Formulierung)	unverständlicher Ausdruck
TRANS DCT Runde 2	117	formuliert kompliziert	unverständlicher Ausdruck
TRANS DCT Runde 3	9	spricht undeutlich	unverständlicher Ausdruck
TRANS DCT Runde 3	14	spricht leise	unverständlicher Ausdruck
TRANS DCT Runde 4	166	spricht schnell	unverständlicher Ausdruck
TRANS DCT Runde 5	30	spricht im Dialekt (teilw. Schwer verständlich)	unverständlicher Ausdruck
TRANS DCT Runde 5	65	zittige Stimme, leise	unverständlicher Ausdruck
TRANS DCT Runde 9	142	muschelt	unverständlicher Ausdruck
TRANS DCT Runde 1	8	spricht klar und deutlich	verständlicher Ausdruck
TRANS DCT Runde 1	67	spricht verständlich	verständlicher Ausdruck
TRANS DCT Runde 2	8	klare und deutliche Stimme	verständlicher Ausdruck
TRANS DCT Runde 2	112	beide reden laut und deutlich	verständlicher Ausdruck
TRANS DCT Runde 3	26	spricht laut und deutlich	verständlicher Ausdruck
TRANS DCT Runde 4	13	spricht ruhig und laut	verständlicher Ausdruck
TRANS DCT Runde 4	150	antwortet kurz	verständlicher Ausdruck
TRANS DCT Runde 6	46	Anweisungen, sehr knapp und präzise	verständlicher Ausdruck
TRANS DCT Runde 3	207	bestätigt Anweisung und fasst zusammen „den 6er auf die 2“	verständlicher Ausdruck
TRANS DCT Runde 3	214	spezifiziert Anweisung von Partner	verständlicher Ausdruck
TRANS DCT Runde 7	80	redet sehr langsam	verständlicher Ausdruck
TRANS DCT Runde 10	127	antwortet (kurz und knapp) 7	verständlicher Ausdruck
TRANS DCT Runde 3	206	muss nachfragen welche E auf welche Str.	versteht Partner nicht
TRANS DCT Runde 3	253	erfragt erneut, ob Beschränkung auf Str. 3	versteht Partner nicht
TRANS DCT Runde 5	167	fragt, ob Ost Vorschlag wiederholen kann	versteht Partner nicht
TRANS DCT Runde 6	110	fragt erneut, ob Str. 4 nur zur Hälfte belastbar	versteht Partner nicht
TRANS DCT Runde 7	211	Missverständnis	versteht Partner nicht
TRANS DCT Runde 1	21	sagt Verhältnisse ausführlich, liest alle Infos ungekürzt vor	gefilterte Infos
TRANS DCT Runde 1	160	rechnet Verhältnis um	gefilterte Infos
TRANS DCT Runde 2	119	berechnet gleich Einschränkung	gefilterte Infos
TRANS DCT Runde 5	88	sagt Einschränkung genau und rechnet sie dann gleich um (im 2. Schritt)	gefilterte Infos
TRANS DCT Runde 6	8	gibt nur wichtige Infos weiter	gefilterte Infos
TRANS DCT Runde 7	171	sagt Belastung, rechnet Beschränkung mit ein	gefilterte Infos
TRANS DCT Runde 10	55	Verhältnisse im 2. Schritt berechnet	gefilterte Infos
TRANS DCT Runde 1	78	Erklärt seine Zuarbeiterrolle	gefilterte Infos
TRANS DCT Runde 1	97	rechnet Beschränkung nicht mit ein	ungefilterte Infos
TRANS DCT Runde 1	157	rechnet Verhältnisse nicht gleich um	ungefilterte Infos
TRANS DCT Runde 1	170	sagt Vorrat (alle Infos Anzahl - Gewicht - Punkte)	ungefilterte Infos
TRANS DCT Runde 1	188	wird nach Belastung einer Str gefragt -antwortet mit gesamten Belastungsstand	ungefilterte Infos
TRANS DCT Runde 4	5	sagt auch Sondermeldung ohne Problem	ungefilterte Infos
TRANS DCT Runde 5	27	sagt Verhältnisse an, ohne sie umzurechnen	ungefilterte Infos

Kognitive Leistung

Video	Absatz	Verhaltensanker	Subkategorie
TRANS_DCT_Runde_5	53	erkennt Dummy	Dummy
TRANS_DCT_Runde_1	89	erfragt Angaben erneut	Merkfehler
TRANS_DCT_Runde_1	148	Erfragt Maximalbelastung erneut	Merkfehler
TRANS_DCT_Runde_1	157	nennt Maximalbelastung: 80-40-30-10 (falsch)	Merkfehler
TRANS_DCT_Runde_2	34	Sagt er habe MaxBel vergessen	Merkfehler
TRANS_DCT_Runde_3	258	erfragt Gewichte von West erneut	Merkfehler
TRANS_DCT_Runde_4	175	fragt wie viel er auf Str. 3 setzen soll (vergessen)	Merkfehler
TRANS_DCT_Runde_3	103	Bittet um 5 E mit möglichst hoher Punktzahl auf Str. 1	Punkteorientiert
TRANS_DCT_Runde_3	260	fragt nach Punktzahl der Gewichte,	Punkteorientiert
TRANS_DCT_Runde_5	145	sagt schwerstes Gewicht mit Punktzahl	Punkteorientiert
TRANS_DCT_Runde_6	89	fragt nach Fahrzeug mit mehr als 5 Punkten	Punkteorientiert
TRANS_DCT_Runde_8	166	zählt Kapazitäten auf und sagt auch Punktwerte	Punkteorientiert
TRANS_DCT_Runde_8	169	sagt aktuellen Punktestand	Punkteorientiert
TRANS_DCT_Runde_1	10	verrechnet sich	Rechenfehler
TRANS_DCT_Runde_7	42	Anweisung, 1 E auf Str. 4 (Verhältnis falsch umgerechnet),	Rechenfehler
TRANS_DCT_Runde_7	143	korrigiert, Str. 4 nur zur Hälfte benutzbar, also 5 E (Beschränkung falsch gerechnet)	Rechenfehler
TRANS_DCT_Runde_1	48	zeit ist noch da	Zeitorientiert
TRANS_DCT_Runde_9	108	Eingabe, fragt nach Zeitangabe bei Anfangsbildschirm	Zeitorientiert
TRANS_DCT_Runde_9	143	Bearbeitungszeit 4 min, will Uhr im Auge behalten	Zeitorientiert
TRANS_DCT_Runde_9	154	sagt wie viel Zeit noch über, ca 10s	Zeitorientiert

QUELLENVERZEICHNIS

- Adlmann, E. M. & Gittler G. (2011). *Persönlichkeits- und Differentielle Psychologie*. Skript zur Vorlesung. Wien: Universität Wien (Fachbereich Psychologie) unveröffentlicht.
- Amelang, M., Bartussek, D. (1981). *Differentielle Psychologie und Persönlichkeitsforschung* (3. Aufl.). Stuttgart: Kohlhammer.
- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6.Aufl.). Stuttgart: Kohlhammer.
- Amelang, M. & Schmidt-Atzert L. (2012): *Psychologische Diagnostik*. Berlin: Springer.
- Amelang, M. & Zielinski, L. (2006). *Psychologische Diagnostik und Intervention* (4. Aufl.). Berlin: Springer.
- Bateson, P. & Martin, P. (1986). *Measuring Behaviour - an introduction guide*. Cambridge: Cambridge University Press.
- Becker-Beck, U. (1997). *Soziale Interaktion in Gruppen – Struktur und Prozessanalyse*. Opladen: Westdeutscher Verlag GmbH.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI)*. Göttingen: Hogrefe.
- Bortz, J. & Döring, N (1995). *Forschungsmethoden und Evaluation* (2. Aufl.). Heidelberg: Springer.
- Caspar, F & Wirtz, M. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

- Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations*, 2, 129-152.
- Deutsch, M. (1976). *Konfliktregelung: konstruktive und destruktive Prozesse*. München: Reinhardt.
- Dollard, J., Doob, L., Miller, N., Mower O. H. & Sears, R. R. (1939). *Frustration and aggression*. New Haven: Yale University Press.
- Eschen-Léguedé, S. (2006). Rater training for an assessment centre in air traffic controller selection: development and evaluation. *Human Factors and Aerospace Safety*, 6, 291-304.
- Eschen-Léguedé, S. (2012). *Beobachtertraining im Kontext von Assessment Centern*. Saarbrücken: AV Akademikerverlag.
- von Eye, A. & Mun, Y.M. (2005). *Analyzing Rater Agreement – Manifest Variable Methods*. Mahwah: Lawrence Erlbaum Associates.
- Faßnacht, G. (1995). *Systemische Verhaltensbeobachtung* (2. Aufl.). München: Ernst Reinhardt, GmbH & Co, Verlag.
- Fleishman E.A., Kleinmann M., Manzey D. & Schumacher S. (2010). *F-JAS - Fleishman Job Analyse System für eigenschaftsbezogene Anforderungsanalysen*. Göttingen: Hogrefe.
- Gerrig, R. & Zimbardo, P. (2008). *Psychologie* (18. Aufl.). München: Pearson Education Deutschland GmbH.
- Gibson, L. W., Lounsbury, J. W., Loveland, J. M., Sundstorm, E. & (2003). Intelligence, „Big Five“ personality traits, and work drive as predictors of course grades. *Personality and Individual Differences*, 35, 1231–1239.
- Graumann, C.F. (1966). Grundzüge der Verhaltensbeobachtung. In E. Meyer (Hrsg.), *Fernsehen in der Lehrerbildung*. München: Manz.

- Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung*. Weinheim: Psychologische Verlagsunion.
- Gros, E. (1994). *Anwendungsbezogene Arbeits-, Betriebs- und Organisationspsychologie. Eine Einführung*. Göttingen: Hogrefe.
- Goodenough, D. R. (1978). Field dependence. In H. London & J. E. Exner (Hrsg.), *Dimension of Personality*. New York: Wiley.
- Harbourn, N. (2007). *Verhalten nach Misserfolg. Extremgruppenvalidierung einer experimentellen Versuchsanordnung zur Erhebung von Frustrationstoleranz*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Hartmann, H. (1970). *Psychologische Diagnostik*. Stuttgart: W. Kohlhammer GmbH.
- Hasemann, K. (1964). *Verhaltensbeobachtung und Verhaltensbeurteilung in der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Heckhausen, H. (1955). Motivationsanalyse der Anspruchsniveau-Setzung. *Psychologische Forschung*, 25, 118-154.
- Hollander, E.P. & Willis, R. H. (1967). Some current issues in the psychology of conformity and nonconformity. *Psychological Bulletin*, 68, 62-76.
- Holocher-Ertl, S. (2008). *Hochbegabungsdagnostik nach dem Wiener Diagnosemodell zum Hochleistungspotenzial*. Dissertation, Universität Wien.
- Holocher-Ertl, S., Kubinger, K.D. & Hohensinn, C. (2006). Zur Definition von Hochbegabung ist die Höhe des IQ zwar Konvention aber völlig ungeeignet: Ein neues Diagnosemodell im Spannungsfeld von Hochbegabung und Hochleistung. In B. Gula, R. Alexandrowicz, S. Strauß, E. Brunner, B. Jenull-Schiefer & O. Vitouch (Hrsg.), *Perspektiven psychologischer Forschung in Österreich. Proceedings*

zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie (S. 444- 451). Lengerich: Pabst.

Hossiep, R. (2003): Assessment Center. In K. D. Kubinger & R. S. Jäger, *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.

Huffcutt, A. I. & Woehr, D. J. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*. 67, 189-205.

Hunter, D. & Martinussen, M. (2008). *Aviation psychology and human factors*. Boca Raton: CRC Press Taylor & Francis Group.

Häcker, H. & Stapf, K.H. (2009). *Dorsch Psychologisches Wörterbuch* (15. Aufl.). Bern: Huber.

Jäger, R. (2003): Arbeitsprobe. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.

Kelava, A. & Moosbrugger, H. (2008). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.

Kleinmann, M. (2003). *Assessment-Center*. Göttingen: Hogrefe.

Kubinger, K. (2009). *Psychologische Diagnostik* (2.Aufl.). Göttingen: Hogrefe.

Kubinger, K.D. & Ebenhö, J. (1997). *Arbeitshaltungen – Kurze Testbatterie: Anspruchsniveau, Frustrationstoleranz, Leistungsmotivation, Impulsivität/Reflexivität*. Frankfurt: Swets.

Litzcke, S. M. & Schuh, H. (2010). *Stress, Mobbing und Burn-out am Arbeitsplatz*. Heidelberg: Springer.

Maltby, J., Day, L. & Macaskill, A. (2011). *Differentielle Psychologie, Persönlichkeit und Intelligenz*. München: Pearson Studium.

- Pecena, Y., Keye, D., Conzelmann, K., Grasshoff, D., Maschke, P., Heintz, A. & Eißfeldt, H. (2013). Predictive Validity of Selection Procedure for Air Traffic Controller Trainees. *Aviation Psychology and Applied Human Factors*, 13, 19-27.
- Rindermann, H. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24, 77-79.
- Scheiderer, J. & Ebermann, H. (2011). *Human Factors in Cockpit*. Heidelberg: Springer.
- Schelling, T.C. (1960). *The strategy of conflict*. Cambridge, Mass.: Harvard University Press.
- Schmidt-Atzert, L., Kersting, M., Preckel F., Westhoff ,K. & Ziegler, M. (2012). Zum Stand der Diagnostik. *Psychologische Rundschau*, 63, 167-174.
- Schrott, A. (2003). Belastbarkeitsdiagnostik. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.
- Schuler, H. (1996). *Psychologische Personalauswahl: Einführung in die Berufsseignungsdiagnostik*. Göttingen: Hogrefe.
- Schulz von Thun, F. (1992). *Miteinander reden: Störungen und Klärungen*. Band 1, Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag GmbH.
- Schwarzer, R. (1981): *Streß, Angst und Hilflosigkeit : die Bedeutung von Kognitionen und Emotionen bei der Regulation von Belastungssituationen*. Stuttgart: Kohlhammer.
- Seiwald, B. (2003): Lügenskala. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.
- Siem, B. & Stürmer, S. (2013). *Sozialpsychologie der Gruppe*. München: Ernst Reinhardt Verlag.

- Stelling, D. (1999). *Teamarbeit in Mensch-Maschine-Systemen* (8. Aufl.). Göttingen: Hogrefe.
- Sonnenberg, H. G., Wottawa, H. (2008) *MAP Management Potenzial Analyse*. Mödling: Schuhfried GmbH.
- Wagner-Menghin, M. (2011). *Arbeitsstil – ein vernachlässigtes Konstrukt in der psychologischen Diagnostik*. Frankfurt am Main: Peter Lang.
- Wieland-Eckelmann, R. (1992). *Kognition, Emotion und psychische Beanspruchung*. Göttingen: Hogrefe.
- Yerkes, R.M. & Dodson, J.D. (1908): The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-48

„Ich habe mich bemüht, sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.“

DANKSAGUNG

Sei auch noch so viel bezeichnet,
Was man fürchtet, was begehrt,
Nur weil es dem Dank sich eignet,
Ist das Leben schätzenswert.

Johann Wolfgang von Goethe

In diesem Sinne danke ich meiner Familie, Denise Bader und Clemens Mielacher für die persönliche Unterstützung. Des Weiteren gilt mein Dank der Bundeswehr für die Genehmigung sowie Prof. Dr. Georg Gittler und Dipl.-Psych. Anne Milatz für die fachlich kompetente Betreuung dieser Arbeit.

EIDESSTAATLICHE VERSICHERUNG

„Ich versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der in den beigefügten Verzeichnissen angegebenen Hilfsmitteln nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Alle Quellen, die dem World Wide Web entnommen oder in einer sonstigen digitalen Form verwendet wurden, sind der Arbeit beigefügt. Der Durchführung einer elektronischen Plagiatsprüfung stimme ich hiermit zu. Die eingereichte elektronische Fassung der Arbeit entspricht der eingereichten schriftlichen Fassung exakt. Die Arbeit wurde bisher keiner anderen Prüfbehörde vorgelegt und auch nicht veröffentlicht. Ich bin mir bewusst, dass eine unwahre Erklärung rechtliche Folgen haben kann.“

Wien, den 27.08.2014

Sarah Baeske

LEBENS LAUF

Name: Sarah Baeske
Wohnort: Beheimgasse 80/5/9, 1170 Wien
Geburtsdatum: 01.08.1990
Geburtsort: Deutschland, Plauen
Staatsangehörigkeit: deutsch

Ausbildung

- Schulausbildung: 1997-2001 Grundschule Reusa
2001-2009 Lessinggymnasium Plauen (Abitur 1,9)
- Studium: 2009-2014 Psychologiestudium Universität Wien
2012 Vordiplom
2014 voraussichtlicher Abschluss des
Diplomstudiums

Berufserfahrung

- Juli 2006 Praktikum Psychiatriepraxis Dr. Juchem (Plauen, Deutschland)
- Juli 2011 Praktikum Kindergarten „Japsnoetjies“ (Lephalale, Südafrika)
- August/ September 2012 Praktikum Personalabteilung „Murray & Roberts“ (Lephalale, Südafrika)
- August/September 2013 Praktikum Bundeswehr, Zentrum für Luft- und Raumfahrtmedizin – Flugpsychologie
- Ab Dezember 2014 Diplomarbeit Bundeswehr, Zentrum für Luft- und Raumfahrtmedizin – Flugpsychologie