



TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Sport- und Gesundheitswissenschaften
Lehrstuhl für Sportpsychologie

Verzerrungstendenzen in neuropsychologischen Untersuchungen:
Eine multimethodale Erfassung und deren diagnostische Grenzen

Stephanie Franziska Stadler

Vollständiger Abdruck der von der Fakultät für Sport- und Gesundheitswissenschaften der
Technischen Universität München zur Erlangung des akademischen Grades eines
Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Joachim Hermsdörfer

Prüfer der Dissertation:

1. Prof. Dr. Jürgen Beckmann

2. apl. Prof. Dr. Thomas Jahn

Die Dissertation wurde am 26.09.2019 bei der Technischen Universität München eingereicht
und durch die Fakultät für Sport- und Gesundheitswissenschaften am 15.01.2021
angenommen.

Danksagung

Zunächst möchte ich mich bei allen Probanden und Patienten bedanken, die an meiner Studie teilgenommen haben. Auf Grund dieser großen Bereitschaft konnte ich meine Arbeit erst durchführen.

Weiter bedanke ich mich ausdrücklich bei Herrn Professor Dr. Thomas Jahn für die Unterstützung und hilfreichen Anregungen. Als Leiter der AG Klinische und Experimentelle Neuropsychologie im Klinikum rechts der Isar der TU München ermöglichte es mir Herr Professor Jahn meine Fragestellungen zu entwickeln und zu operationalisieren. Zudem danke ich Herrn Professor Dr. Jürgen Beckmann für die Möglichkeit an der Fakultät für Sport- und Gesundheitswissenschaften zu promovieren.

Außerdem danke ich allen, die mich durch Beratung, Korrekturen und Rekrutierung von Studienteilnehmern unterstützt haben: Andreas Böck, Lena Stadler, Richard Stadler, Anna Nützel, Denise Wenning und die Statistikberatung der TU München. Meiner Familie und meinen Freunden danke ich für den Zuspruch und die aufbauenden Worte, die mir geholfen haben, die Arbeit zu Ende zu schreiben.

Zudem bedanke ich mich bei Herrn Dr. Christoph Siebold dafür, dass ich im Neurozentrum Starnberg Patienten rekrutieren konnte.

Mein besonderer Dank gilt der Frauenförderung der TU München. Durch das Förderprogramm „Chancengleichheit für Frauen in Forschung und Lehre“ konnte ich meine Arbeit größtenteils abschließen.

Inhaltsverzeichnis

Danksagung.....	II
Inhaltsverzeichnis.....	III
Tabellenverzeichnis.....	VIII
Abbildungsverzeichnis.....	XI
Abkürzungsverzeichnis.....	XII
Zusammenfassung.....	XV
1 Einleitung.....	1
2 Beschwerdvalidierung in der Neuropsychologie.....	6
2.1 Begriffsbestimmung.....	6
2.2 Antwortverzerrungen in diversen Kontexten.....	8
2.2.1 Simulation und Aggravation.....	9
2.2.2 Artificielle Störung.....	10
2.2.3 Somatoforme und dissoziative Störungen.....	11
2.2.4 Differentialdiagnostische Unterscheidung.....	12
2.3 Diagnostische Kriterien der Simulation kognitiver Beeinträchtigungen.....	13
2.3.1 Diagnostische Kriterien nach Slick.....	14
2.3.2 Modifikation und Erweiterung der Kriterien nach Slick.....	17
2.4 Studiendesigns der BV.....	18
2.5 Prävalenzraten.....	20
2.5.1 Kontextspezifische Prävalenzen.....	20
2.5.2 Diagnosespezifische Prävalenzen.....	22
2.6 Erklärungsmodelle, kognitive Prozesse und neuronale Korrelate.....	23
2.6.1 Erklärungsmodelle nach Rogers.....	24
2.6.2 Anreize als Erklärung.....	25
2.6.3 Kognitive und neuronale Prozesse des Lügens.....	26
2.6.4 Neuronale Korrelate in bildgebenden Verfahren.....	27
2.7 Negative Antwortverzerrungen in neuropsychologischen Untersuchungen.....	29
2.7.1 Beeinflussbarkeit negativer Antwortverzerrungen.....	29
2.7.2 Erfassung negativer Antwortverzerrungen.....	31

2.7.3	Diagnostische Ansätze.....	31
2.8	Performanzvaliditätstests (PVT)	36
2.8.1	Konstruktionsprinzipien	37
2.8.2	Gütekriterien der PVT	40
2.8.3	Interpretation der PVT	42
2.8.4	Word Memory Test (WMT).....	44
2.8.5	Amsterdamer Kurzzeitgedächtnistest (AKGT)	48
2.8.6	Aggravations- und Simulations-Test (AST).....	51
2.9	Analyse neuropsychologischer Testprofile	52
2.9.1	Eingebettete Parameter des California Verbal Learning Test (CVLT)	54
2.9.2	Reaktionszeiten als eingebettete Parameter.....	60
2.10	Diagnostische Güte und Anzahl verwendeter Methoden.....	64
2.11	Diagnostische Grenzen der Beschwerdvalidierung.....	65
2.11.1	Beschwerdvalidierung und Coaching	65
2.11.2	Beschwerdvalidierung und kognitive Dysfunktionen.....	67
3	Fragestellungen und Hypothesen	72
3.1	Erster Teil der Studie: Testperformanz im experimentellen Analogdesign.....	74
3.2	Zweiter Teil der Studie: Testperformanz der Patienten mit negativen Antwortverzerrungen (PNA).....	75
3.3	Dritter Teil der Studie: Testperformanz der Patienten mit kognitiven Beeinträchtigungen (PKB)	75
4	Methodik: erster Teil der Studie	76
4.1	Studiendesign und Ablauf.....	76
4.1.1	Instruktionen und Erinnerungsankündigung	78
4.1.2	Erhebung psychosozialer und relevanter experimenteller Einflussvariablen...	79
4.2	Untersuchungsverfahren	81
4.2.1	Neuropsychologische Leistungstests	81
4.2.2	Additionalen Leistungstests.....	86
4.2.3	Performanzvalidierungstests (PVT)	88
4.3	Fallzahlberechnung	92

4.3.1	Ein- und Ausschlusskriterien.....	95
4.3.2	Rekrutierung und Untersuchungsbedingungen	95
4.3.3	Beschreibung der Stichprobe der Analogstudie	96
4.4	Statistische Datenanalysen	99
4.4.1	Datenaufbereitung	99
4.4.2	Methodenauswahl und Variablenselektion.....	101
5	Methodik: zweiter Teil der Studie.....	106
5.1	Untersuchungsbedingungen	106
5.2	Beschreibung der Stichprobe PNA	107
5.3	Statistische Datenanalysen	110
6	Methodik: dritter Teil der Studie.....	111
6.1	Studiendesign und Ablauf.....	111
6.2	Untersuchungsverfahren	111
6.3	Soziodemographische und klinische Daten der PKB.....	114
6.4	Statistische Datenanalysen	116
7	Ergebnisse: erster Teil der Studie	118
7.1	Deskriptive Ergebnisdarstellung	118
7.2	Inferenzstatistische Gruppenunterschiede im Analogdesign	120
7.2.1	Testperformanz in den Performanzvalidierungstests (PVT)	121
7.2.2	Testperformanz in neuropsychologischen Standardverfahren	121
7.2.3	Ergebnisse zu Fragestellung F1 und Hypothese H1 und H2	125
7.3	Einfluss der Aufgabenschwierigkeit	126
7.3.1	Selbstbeurteilungen der subjektiven Schwierigkeit.....	126
7.3.2	Ergebnisse zu Fragestellung F2 und Hypothese H3	130
7.4	Klassifikation anhand von Performanzvaliditätstests	130
7.4.1	Klassifikation anhand verfahrensspezifischer Schwellenwerte.....	130
7.4.2	Statistische Klassifikation der Gruppenzugehörigkeit	131
7.4.3	Ermittlung empirischer Schwellenwerte	133
7.4.4	Ergebnisse zu Fragestellung F3 und Hypothese H4.....	134
7.5	Klassifikation anhand neuropsychologischer Standardverfahren	134

7.5.1	Klassifikation anhand von CVLT-Kennwerten.....	134
7.5.2	Ergebnisse zu Fragestellung F4 und Hypothese H5.....	138
7.5.3	Klassifikation anhand von TAP-Kennwerten.....	138
7.5.4	Kombination der TAP- und CVLT-Kennwerte.....	145
7.5.5	Ergebnisse zu Fragestellung F4 und Hypothese H6.....	147
8	Ergebnisse: zweiter Teil der Studie.....	148
8.1	Deskriptive Ergebnisdarstellung.....	148
8.2	Performanzvaliditätstests.....	150
8.2.1	Inferenzstatistische Gruppenunterschiede.....	150
8.2.2	Klassifikation mit dem AST.....	151
8.2.3	Ergebnisse zu Fragestellung F5 und Hypothese H7.....	151
8.3	Neuropsychologische Standardverfahren.....	152
8.3.1	Inferenzstatistische Gruppenunterschiede.....	152
8.3.2	Klassifikationen mit dem CVLT und der TAP.....	153
8.3.3	Ergebnisse zu Fragestellung F5 und Hypothese H8.....	154
9	Ergebnisse: dritter Teil der Studie.....	156
9.1	Performanzvalidierungstests.....	156
9.1.1	Ergebnisse zu Fragestellung F6 und Hypothese H9.....	157
9.2	Kognitive Funktionen und PVT.....	158
9.2.1	Ergebnisse zu Fragestellung F7 und Hypothese H10.....	160
10	Diskussion.....	161
10.1	Zusammenfassung der Ergebnisse.....	161
10.2	Generelle Einschränkungen der Studie.....	163
10.3	Diskussion Hypothese H1.....	167
10.4	Diskussion Hypothese H2.....	168
10.5	Diskussion Hypothese H3.....	169
10.6	Diskussion Hypothese H4.....	170
10.7	Diskussion Hypothese H5.....	171
10.8	Diskussion Hypothese H6.....	172
10.9	Diskussion Hypothese H7.....	174

10.10 Diskussion Hypothese H8	175
10.11 Diskussion Hypothese H9 und H 10	177
11 Implikationen für die Praxis und Ausblick	180
Literaturverzeichnis	181
Anhang	199

Tabellenverzeichnis

Tabelle 1	<i>Begriffe und Definitionen der BV.</i>	7
Tabelle 2	<i>Klassifikation nicht-authentischer Störungen anhand Bewusstseinsgrad und Motivation (modifiziert nach Merten (2011) und Slick et al. (1999)).</i>	13
Tabelle 3	<i>Beschreibung der Kriterien nach Slick et al. (1999).</i>	15
Tabelle 4	<i>Adjustierte Prävalenzraten wahrscheinlicher Simulation / Aggravation einzelner Diagnosegruppen nach Mittenberg et al. (2002).</i>	23
Tabelle 5	<i>Modifizierte Kurzbeschreibung deutschsprachiger PVT nach Merten (2011, 2014) und deren Einsatzhäufigkeit laut Dandachi-FitzGerald et al. (2015).</i>	39
Tabelle 6	<i>Klassifikationsstatistiken der PVT in den Metaanalysen von Vickery et al. (2001) und Sollman & Berry (2011).</i>	42
Tabelle 7	<i>Stichprobenumfangsplanung aus Effektstärken der Forschungsliteratur.</i>	94
Tabelle 8	<i>Deskriptive Statistiken soziodemographischer Daten der Gesamtstichprobe und getrennt nach randomisierter Gruppenzuweisung.</i> ...	97
Tabelle 9	<i>Deskriptive Statistiken experimenteller Einflussvariablen.</i>	98
Tabelle 10	<i>Deskriptive Statistiken der Anstrengungs- und Simulationsbereitschaft (fünfstufige Einschätzungen von 0 „kaum vorhanden“ bis 4 „sehr stark vorhanden“).</i>	99
Tabelle 11	<i>Deskriptive Statistiken soziodemographischer Daten der PNA und ES, deren Gesamtstichprobe und Ergebnisse des statistischen Vergleichs der Teilstichproben.</i>	109
Tabelle 12	<i>Deskriptive Statistiken soziodemographischer und klinischer Daten der PNA und PKB, deren Gesamtstichprobe und Ergebnisse des statistischen Vergleichs der Teilstichproben.</i>	115
Tabelle 13	<i>Deskriptive Statistiken und Effektstärken der PVT-Gruppenvergleiche.</i>	121
Tabelle 14	<i>CVLT: Deskriptive Statistiken und Effektstärken der Mittelwertdifferenzen.</i>	121
Tabelle 15	<i>TAP: Deskriptive Statistiken und Effektstärken der Mittelwertdifferenzen.</i> .	122
Tabelle 16	<i>Additionalen Leistungstests: Deskriptive Statistiken und Effektstärken der Mittelwertdifferenzen.</i>	123

Tabelle 17	<i>Subjektiven Schwierigkeit: Deskriptive Statistiken, Ergebnisse der t-Tests und Effektstärken.</i>	127
Tabelle 18	<i>Paarweise Vergleiche der subjektiven Schwierigkeiten der PVT mittels t-Tests für verbundene Stichproben separat für die KG und ES.</i>	128
Tabelle 19	<i>Korrelationskoeffizienten nach Pearson der ES (n = 33) der Testperformanz im AKGT, AST, WMT und der subjektiven Schwierigkeitsbeurteilungen.</i>	128
Tabelle 20	<i>Vorhersage der PVT-Testleistungen durch die subjektive Schwierigkeit, Gruppenzugehörigkeit und deren Wechselwirkung (Schwierigkeit x Gruppe) in der Gesamtstichprobe (N = 66) anhand blockweiser multipler Regressionsanalysen.</i>	129
Tabelle 21	<i>Klassifikationsstatistiken gemäß verfahrensspezifischer Schwellenwerte.</i>	130
Tabelle 22	<i>Minimale und maximale Testperformanz der einzelnen PVT-Kennwerte.</i>	131
Tabelle 23	<i>Separate binär-logistische Regressionsanalysen jeweils mit dem AST und WMT IR.</i>	132
Tabelle 24	<i>ROC-Kurven der PVT: AUC und Wilcoxon-Teststatistiken.</i>	133
Tabelle 25	<i>Klassifikationsstatistiken des AGKT, AST und WMT.</i>	133
Tabelle 26	<i>CVLT: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.</i>	135
Tabelle 27	<i>ROC-Kurven des CVLT: AUC und Wilcoxon-Teststatistiken.</i>	136
Tabelle 28	<i>Klassifikationsstatistiken der CVLT-Parameter.</i>	137
Tabelle 29	<i>TAP Alertness: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.</i>	139
Tabelle 30	<i>TAP Geteilte Aufmerksamkeit: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.</i>	140
Tabelle 31	<i>TAP Go/Nogo: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.</i>	141
Tabelle 32	<i>ROC-Kurven der TAP-Kennwerte: AUC und Wilcoxon-Teststatistiken.</i>	142
Tabelle 33	<i>Klassifikationsstatistiken der TAP-Kennwerte.</i>	144
Tabelle 34	<i>TAP und CVLT: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.</i>	145
Tabelle 35	<i>Kennwerte der Modellgüte und Modellanpassung der logistischen Regressionsmodelle.</i>	146

Tabelle 36	<i>Deskriptive Statistiken und Effektstärkenmaße der PVT-Gruppenvergleiche.</i>	150
Tabelle 37	<i>Klassifikationen der ES und PNA mit AST-Schwellenwerten.</i>	151
Tabelle 38	<i>Deskriptive Statistiken und Effektstärken der Gruppenunterschiede im CVLT.</i>	152
Tabelle 39	<i>Deskriptive Statistiken und Effektstärken der Gruppenunterschiede in den TAP-Untertests.</i>	153
Tabelle 40	<i>Klassifikationen der ES und PNA mit CVLT und TAP-Kennwerten.</i>	154
Tabelle 41	<i>Deskriptive Statistiken und Effektstärken der PVT-Gruppenvergleiche.</i>	156
Tabelle 42	<i>Klassifikationsstatistiken der AKGT- und AST-Parameter.</i>	157
Tabelle 43	<i>Deskriptive Statistiken: z-standardisierte Parameter der CERADplus-NTB der PKB (n = 12).</i>	158
Tabelle 44	<i>Korrelationskoeffizienten nach Spearman der PKB (n = 12) des AKGT, AST und der CERADplus-NTB.</i>	159

Abbildungsverzeichnis

- Abbildung 1** Differenzierung nicht-authentischer Störungen nach Heubrock et al. (2013). 8
- Abbildung 2** Darstellung der sechs Testfolgen..... 77
- Abbildung 3** Mittlere Prozentwerte mit 95%-Konfidenzintervallen der KG ($n = 33$) und ES ($n = 33$) im *Amsterdamer Kurzzeitgedächtnistest (AKGT)*, *Aggravations- und Simulationstests (AST)* und *Word Memory Test (WMT): unmittelbares Wiedererkennen (WMT IR)*, *verzögertes Wiedererkennen (WMT DR)* und *Konsistenzmaß (WMT Con)*. 118
- Abbildung 4** Mittlere *T*-Werte mit 95%-Konfidenzintervallen der KG ($n = 33$) und ES ($n = 33$): *California Verbal Learning Test (CVLT)*, *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* und *additionalen Leistungstests*. Graue Fläche markiert Normbereich ($40 \leq T \leq 60$). 119
- Abbildung 5** Cohens *d* als Maß der Effektstärke in den PVT und den neuropsychologischen Standardverfahren. 124
- Abbildung 6** Mittelwerte mit 95%-Konfidenzintervallen der KG ($n = 33$) und ES ($n = 33$) der subjektiven Schwierigkeit (fünfstufige Einschätzung von 0 „sehr einfach“ bis 5 „sehr schwer“) im *Amsterdamer Kurzzeitgedächtnistest (AKGT)*, *Aggravations- und Simulationstests (AST)*, *Word Memory Test (WMT)*, *California Verbal Learning Test (CVLT)*, *Rey Complex Figure Test and Recognition Trial (RCFT)*, *Turm von London (TL-D)*, *Zahlen-Verbindungs-Test (ZVT)* und der *Testbatterie zur Aufmerksamkeitsprüfung (TAP): Alertness (TAP:A)*, *geteilte Aufmerksamkeit (TAP: GA)*, *GoNogo (TAP:GN)*. 126
- Abbildung 7** Mittlere Prozentwerte mit 95%-Konfidenzintervallen der PNA ($n = 10$) und ES ($n = 33$) im *Amsterdamer Kurzzeitgedächtnistest (AKGT)*, *Aggravations- und Simulationstests (AST)*, *Word Memory Test (WMT): unmittelbares Wiedererkennen (WMT IR)*, *verzögertes Wiedererkennen (WMT DR)* und *Konsistenzmaß (WMT Con)*. 148
- Abbildung 8** Mittlere *T*-Werte mit 95%-Konfidenzintervallen der PNA ($n = 10$) und ES ($n = 33$): *California Verbal Learning Test (CVLT)* und *Testbatterie zur Aufmerksamkeitsprüfung (TAP)*. Graue Fläche markiert Normbereich ($40 \leq T \leq 60$). 149

Abkürzungsverzeichnis

AKGT	Amsterdam Kurzzeit Gedächtnistest
APA	American Psychological Association
AST	Aggravations- und Simulationstest
ASTM	Amsterdam Short-Term Memory Test
AUC	Area under the curve
BV	Beschwerdenvalidierung
BVT	Beschwerdenvalidierungstest
CASMIN	Comparative Analysis of Social Mobility in Industrial Nations
CARB	Computerized Assessment of Response Bias
CERADplus- NTB	Consortium to Establish a Registry for Alzheimer's Disease Plus NTB
CI	Confidence Interval (Konfidenzintervall)
CS	Coaching Simulanten (speziell instruierte experimentelle Simulanten)
CVLT	California Verbal Learning Test
CVLT-I	California Verbal Learning Test, erste englischsprachige Version
CVLT-II	California Verbal Learning Test, zweite englischsprachige Version
CTIP	Computerized Tests of Information Processing
<i>d</i>	Effektstärke nach Cohen
DCT	Dot Counting Test
DMT	Digit Memory Test
DGPPN	Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde
DRT	Digit Recognition Test
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders in der vierten Auflage mit Textrevision
ES	Experimentelle Simulanten
FIT	15-Item Memory Test
fMRT	funktionelle Magnetresonanztomographie
GDS-15	Geriatrische Depressionsskala-15
GMIP	Genuine Memory Impairment Profile (Profil tatsächlicher Gedächtnisstörungen)

ICD-10	International Statistical Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme) in der zehnten Auflage
ICD-10-GM	International Statistical Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme) in der zehnten Auflage, German Modification
KG	Kontrollgruppe
LMT	Letter Memory Test
<i>M</i>	Mittelwert
MCI	Mild Cognitive Impairment
<i>Md</i>	Median
MMPI	Minnesota Multiphasic Personality Inventory
MMPI-2	Minnesota Multiphasic Personality Inventory-2
MMSE	Mini-Mental-Status Exam
MSVT	Medical Symptom Validity Test
<i>OR</i>	Odds Ratio (Quotenverhältnis)
PDRT	Portland Digit Recognition Test
PET	Positronen-Emissions-Tomographie
PKB	Patienten mit kognitiven Beeinträchtigungen (PKB)
PNA	Patienten mit negativen Antwortverzerrungen
PTBS	Posttraumatische Belastungsstörung
PVT	Performanzvaliditätstests
<i>r</i>	Produkt-Moment-Korrelation
RCFT	Rey-Osterrieth Complex Figure Test
ROC	Receiver-Operating-Characteristic
RDS	Reliable Digit Span
RZ	Reaktionszeit
<i>SD</i>	Standard Deviation (Standardabweichung)
SFSS	Strukturierten Fragebogen Simulierter Symptome
SHT	Schädel-Hirn-Trauma
SIMS	Structured Inventory of Malingered Symptomatology
SIRS	Structured Interview of Reported Symptoms

SND	Simulation neurokognitiver Dysfunktionen
SRSI	Self-Report Symptom Inventory
SRVT	Selbstauskunftsvaliditätstest
TAP	Testbatterie zur Aufmerksamkeitsprüfung
TBFN	Testbatterie für Forensische Neuropsychologie
TMT-A	Trail-Making-Test Teil A
TMT-B	Trail-Making-Test Teil B
TL-D	Turm von London in der deutschen Version
TOMM	Test of Memory Malingering
WRG	Wiener Reaktionsgerät
WMS-III	Wechsler Memory Scale-III
WMT	Word Memory Test
VSVT	Victoria Symptom Validity Test
ZVT	Zahlen-Verbindungs-Test

Zusammenfassung

Ziel: Die Überprüfung der Glaubhaftigkeit kognitiver Funktionseinbußen findet in den letzten Jahren sowohl in der neuropsychologischen Praxis als auch in Forschungsarbeiten vermehrt Beachtung. In der vorliegenden Arbeit wurden verschiedene diagnostische Verfahren zur Erfassung verzerrter kognitiver Symptomdarstellungen miteinander verglichen. Neben den dafür konstruierten Performanzvalidierungstests (PVT) wurden auch neuropsychologische Standardverfahren analysiert.

Methode: In einer Analogstudie mit gesunden Probanden (KG, $n = 33$) und experimentellen Simulanten (ES, $n = 33$) wurden die Leistungsdifferenzen zwischen den Gruppen ermittelt. Die Anstrengungsbereitschaft wurde mit drei PVT erfasst: *Word Memory Test (WMT: Green, 2003, Revised 2005)*, *Amsterdamer Kurzzeitgedächtnistests (AKGT: Schmand, Lindeboom, Merten & Millis, 2005b)* und *Aggravations- und Simulationstests (AST: Eberl & Wilhelm, 2007)*. Zudem wurde der Zusammenhang zwischen der subjektiven Aufgabenschwierigkeit und der präsentierten Testperformanz untersucht.

Darüber hinaus wurde die experimentelle Gruppenzugehörigkeit aus den Kennwerten des *California Verbal Learning Test (CVLT: Niemann, Sturm, Thöne-Otto & Willmes, 2008)* und der *Testbatterie zur Aufmerksamkeitsprüfung (TAP: Zimmermann & Fimm, 2012)* durch logistische Regressionsanalysen bestimmt. In ROC-Analysen wurden optimale Schwellenwerte ermittelt.

Anschließend wurden die Ergebnisse der Analogstudie mit einer psychiatrisch-neurologischen Patientenstichprobe mit negativen Antwortverzerrungen (PNA, $n = 10$) verglichen. Mit einer weiteren Patientenstichprobe mit kognitiven Beeinträchtigungen (PKB, $n = 12$) wurden die diagnostischen Grenzen des *AKGT* und *AST* untersucht und die Korrelationen mit den kognitiven Funktionsbereichen der *CERADplus-NTB* berechnet, die bei der Bearbeitung dieser PVT beansprucht werden.

Ergebnisse: Insgesamt präsentierten die ES sowohl in den PVT als auch in den neuropsychologischen Standardverfahren eine schlechtere Testperformanz als die KG. Zudem schätzten beide Experimentalgruppen den *AKGT* und *WMT* schwieriger als den *AST* ein. Es wurde jedoch kein bedeutsamer Zusammenhang zwischen Schwierigkeit und Testperformanz ermittelt.

Mit einer geringeren Sensitivität lagen die Klassifikationsstatistiken des *AST* unterhalb des *WMT* und *AKGT*, die jedoch durch einen höheren Schwellenwert des *AST* (95.7% korrekte

Antworten) verbessert werden konnte. Des Weiteren konnten mit einer Kombination aus jeweils zwei *CVLT*-Kennwerten 98.5% der experimentellen Probanden vorhergesagt werden, aus zwei *TAP*-Parametern 93.9%. Ferner lag die Klassifikationsgenauigkeit mit einer Kombination aus *CVLT*- und *TAP*-Kennwerten bei 98.5%.

In weiteren Analysen wurden die Schwellenwerte und Vorhersagemodelle in der Patientenstichprobe mit negativen Antwortverzerrungen PNA überprüft. Auch in der PNA-Stichprobe konnten durch Kombinationen aus *TAP*- und *CVLT*-Kennwerten 90% bis 100% der PNA als Richtig-Positive klassifiziert werden. Dagegen konnten die hohen Klassifikationsstatistiken des *AST* in der PNA-Stichprobe nicht repliziert werden und es wurden lediglich Richtig-Positiv-Klassifikationen von 30% (Schwellenwert 90%) bzw. 40% (Schwellenwert 95.7%) ermittelt.

Der Vergleich der beiden klinischen Stichproben PNA und PKB verdeutlichte, dass bis zu 75% der Patienten mit kognitiven Beeinträchtigungen der PKB-Stichprobe im *AST* und *AKGT* als Falsch-Positive fehlklassifiziert wurden. Beide PVT korrelierten hoch mit der kognitiven Informationsverarbeitungsgeschwindigkeit. Ferner zeigten sich statistisch bedeutsame Zusammenhänge zwischen dem *AST* und dem *AKGT* und der Wortflüssigkeit der exekutiven Funktionen. Darüber hinaus korrelierte der *AST* mit den visuokonstruktiven Fähigkeiten und verbal-mnestischen Funktionen, der *AKGT* mit den figuralen Gedächtnisleistungen.

Schlussfolgerung: Die Ergebnisse verdeutlichten eine geringe Sensitivität des *AST* im klinischen Kontext. Dagegen stellte insbesondere die Kombination aus *CVLT*- und *TAP*-Parametern eine vielversprechende diagnostische Strategie dar. Zudem wurden diverse Zusammenhänge zwischen den kognitiven Funktionen und der Bearbeitung des *AST* und *AKGT* ermittelt.

1 Einleitung

Es war einmal ein Hirtenjunge, der hütete jeden Tag die Schafe. Morgens nahm er die Tiere von ihren Besitzern in seine Obhut und trieb sie zu satten Wiesen, wo sie weiden konnten. Und zum Abend hin brachte er die Schafe zurück ins Dorf.

Eines Tages, als er den Schafen wie jeden Tag beim Weiden zusah, war dem Hirtenjungen furchtbar langweilig. Da ersann er einen Streich. Er lief in Richtung des Dorfes und rief, so laut er konnte: „Der Wolf kommt! Der Wolf! Er kommt und will sich ein Schaf reißen!“

Sogleich eilten die Leute aus dem Dorf herbei, um mit Knüppeln und Mistgabeln bewaffnet den Wolf zu verjagen. Doch als sie ankamen, war da kein Wolf. Der Hirtenjunge aber brach vor Lachen in Tränen aus und kugelte sich zwischen den weidenden Schafen. Die Herbeigeeilten trotteten murrend zurück ins Dorf.

Einige Tage später langweilte sich der Hirtenjunge erneut. Da fiel ihm sein gelungener Streich ein. Was einmal gelingt, das wird auch ein zweites Mal gelingen, dachte er bei sich. Und wieder rief er: „Der Wolf kommt! Der Wolf! Er kommt und will sich ein Schaf reißen!“ Und wieder eilten die Dorfbewohner herbei, um den Wolf zu verjagen. Und wieder war weit und breit kein Wolf zu sehen.

Eines Abends im Herbst machte sich der Hirtenjunge soeben mit seinen Schafen auf dem Weg zum Dorf. Da sprang ein Wolf aus dem Dickicht und baute sich vor ihnen auf. Der Hirtenjunge bekam es mit der Angst zu tun, schrie aus vollem Halse: „Der Wolf kommt! Der Wolf! Er kommt und will sich ein Schaf reißen!“ Aber kein einziger Dorfbewohner eilte herbei, um dem Jungen zu helfen.

Und so trieb der Wolf sämtliche Schafe in den Wald, wo er sie eines nach dem anderen auffraß. (Aesop, zitiert nach Grabowsky, 2016, S. 109)

Wie der Hirtenjunge in Aesops Fabel können auch Patienten durch Täuschung und Irreführung ihre Glaubwürdigkeit verlieren und die notwendige Diagnostik und Behandlung der zugrunde liegenden Pathologie erschweren oder gar verhindern. Vor allem die Therapie psychischer Störungen basiert in hohem Maße auf den Selbstauskünften und Beschwerdendarstellungen des Patienten (Schmidt, Lanquillon & Ullmann, 2011) und erfordert die Beurteilung nicht unmittelbar beobachtbarer, schwer objektivierbarer innerer Vorgänge. Somit ist die Diagnostik psychischer Störungen stark durch das Verhalten und die Absichten des Patienten

beeinflussbar. Dies bezieht sich auch auf die Darstellung kognitiver Störungen, die subjektiv bspw. durch Einschränkungen der Konzentrations- oder Merkfähigkeit im Alltag wahrgenommen werden. Sie werden mittels verminderter Leistungen in psychometrischen Testverfahren verifiziert und gelten auf Grund der leichten Manipulierbarkeit als besonders anfällig für Verzerrungs- und Täuschungsabsichten (Merten, 2009). Ein Patient braucht eine Aufgabe lediglich ein wenig langsamer zu bearbeiten, etwas weniger zu erinnern oder insgesamt weniger Aufgaben zu lösen (Heubrock & Petermann, 2000). Als Voraussetzung für die Diagnostik kognitiver Störungen gilt demnach die Anstrengungsbereitschaft, während der Bearbeitung der Testverfahren die bestmögliche Leistung zu zeigen (Merten & Dohrenbusch, 2010). Da die Anstrengungsbereitschaft dem Diagnostiker meist verborgen bleibt und mit einer diagnostischen Unsicherheit einhergeht, bezeichnet Bigler (2015) sie als Achillesferse neuropsychologischer Untersuchungen.

Insgesamt stellt die Unterscheidung von vorgetäuschten und tatsächlichen Beschwerden eine Herausforderung im klinischen Alltag dar. So gelang es bspw. in einer Studie von Heaton, Smith, Lehman und Vogt (1978) selbst erfahrenen Klinikern nicht besser als dem Zufall, zwischen Personen mit vorgetäuschten kognitiven Beeinträchtigungen und Patienten mit schweren traumatischen Hirnverletzungen zu unterscheiden. Auch die bekannte Studie von Rosenhan verdeutlichte schon 1973 eindrucksvoll die Schwierigkeiten, vorgetäuschte Symptome zu erkennen. In dieser Untersuchung berichteten Pseudopatienten im Aufnahmegespräch einer stationären Behandlung, unter akustischen Halluzinationen zu leiden. Im Gegensatz zu den Mitpatienten erkannte keiner der Behandler die fehlende zugrundeliegende Pathologie der initialen Symptompräsentation. Als größter Stolperstein in der Differenzierung zwischen authentischen und nicht authentischen Beschwerden wird das Vertrauen eines Klinikers in die eigene Urteilsgüte betrachtet (Merten, 2010). Dabei unterliegen Behandler einem Bias hinsichtlich Vertrauen und Empathie für ihre Patienten, was durch eine entsprechende selektive Informationssuche bestätigt wird und wiederum den Bias in die eigene Urteilsgüte verstärkt (Westhoff & Kluck, 2008). Die Schwierigkeiten, vorgetäuschte Symptome zu erkennen, verdeutlichen die Notwendigkeit den diagnostischen Prozess durch standardisierte und objektivierbare Methoden zu ergänzen.

Bereits in den 60er Jahren befassten sich Rey (1964) sowie Benton und Spreen (Benton & Spreen, 1961; Spreen & Benton, 1963) mit vorgetäuschten kognitiven Beschwerden. Der systematischen Überprüfung der Glaubhaftigkeit und Authentizität demonstrierter kognitiver Beeinträchtigungen wird allerdings erst seit den 1980er Jahren Interesse geschenkt (Merten,

2014) und gelangte nur langsam in den Fokus der neuropsychologischen Forschung. Mit der differenzierten Formulierung diagnostischer Kriterien für das Simulieren kognitiver Beeinträchtigungen legten Slick, Sherman und Iverson (1999) den Grundstein der neuropsychologischen *Beschwerdvalidierung (BV)*. Daraufhin erfolgte ein regelrechter Aufschwung an empirischen Untersuchungen und konzeptionellen Ansätzen der BV und Täuschungsabsichten kognitiver Symptome rückten vermehrt in das neuropsychologische Interesse (Chafetz et al., 2015; Martin, Schroeder & Odland, 2015; Merten, 2002). Besonderes Aufsehen erregte die Studie von Green, Rohling, Lees-Haley und Allen (2001), in der die Anstrengungsbereitschaft einen größeren Einfluss auf die Testleistung hatte, als das tatsächliche Ausmaß kognitiver Beeinträchtigungen. Die Anstrengungsbereitschaftklärte 53% der Varianz innerhalb der neuropsychologischen Testergebnisse auf. Constantinou, Bauer, Ashendorf, Fisher und McCaffrey (2005) replizierten dieses Ergebnis mit weiteren neuropsychologischen Testverfahren. Die Autoren konnten 47% der Varianz auf die Anstrengungsbereitschaft zurückführen und die Relevanz der BV in neuropsychologischen Untersuchungen untermauern.

2005 fasste die *National Academy of Neuropsychology (NAN)* die wichtigsten Gesichtspunkte einer verfahrensübergreifenden, ganzheitlichen Erfassung nicht-authentischen Testverhaltens zusammen. Darin forderten Bush et al. (2005) BV als grundlegenden Bestandteil einer neuropsychologischen Untersuchung zu betrachten. Zudem bezeichneten sie den Einsatz spezieller Methoden zur Erfassung der Anstrengungsbereitschaft bei entsprechendem Anlass als medizinische Notwendigkeit. Des Weiteren wurden 2008 die diagnostischen Vorgehensweisen in der Konsensuskonferenz der *American Academy of Clinical Neuropsychology (AACN)* diskutiert und Empfehlungen für die neuropsychologische Praxis aus der empirischen Datenlage abgeleitet (Heilbronner, Sweet, Morgan, Larrabee & Millis, 2009). BV wurde im forensischen Kontext als medizinisch erforderlich und im klinischen Kontext als wichtiger Beitrag zum Verständnis klinischer Symptome bezeichnet und in der klinischen Alltagspraxis empfohlen. Auch die *Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)* betont in einer Stellungnahme von 2011 vor allem im gutachterlichen Kontext eine Überprüfung der BV, insbesondere wenn von einer Diagnose Kompensationszahlungen abhängen (Dressing, Foerster, Widder, Schneider & Falkai, 2011).

Im deutschsprachigen Raum findet BV hauptsächlich im forensischen und gutachterlichen Kontext Beachtung (Merten, 2009). In der psychiatrischen Praxis begegnet man der

standardisierten BV noch vermehrt mit Skepsis (Merten et al., 2013). Während im forensischen Kontext 45% der deutschsprachigen Neuropsychologen standardisierte BV-Verfahren verwenden, liegt die Häufigkeit im klinischen Alltag bei 8% (Dandachi-FitzGerald, Merten, Ponds & Niemann, 2015). Insgesamt ist in den letzten Jahren neben einer steigenden Anzahl verschiedener BV-Verfahren und einem wachsenden Forschungsinteresse auch eine breitere Akzeptanz in der klinischen Praxis zu verzeichnen. Zudem findet die BV vermehrt Beachtung in der Rechtsprechung, da bereits in einzelnen Gerichtsurteilen die BV präsentierter Symptome berücksichtigt wurde (Merten, 2011).

Der zunehmende Einsatz von BV-Verfahren geht auch mit einer anhaltenden Diskussion über die Interpretation der Ergebnisse einher. Laut dem 2015 veröffentlichten Bericht des unabhängigen *Institute of Medicine (IOM)* liegt im Bereich der BV ein genereller Mangel an empirischen Arbeiten vor sowie eine fehlende Validierung der Gütekriterien einzelner BV-Verfahren, Schwierigkeiten bei der Bestimmung zuverlässiger Prävalenzraten vorgetäuschter Störungen und eine mangelnde Generalisierbarkeit zahlreicher Studien (Freedman & Manly, 2015). Dennoch empfiehlt die *American Academy of Clinical Neuropsychology (AACN)* in einem aktuellen Positionsbericht einen umsichtigen Einsatz von BV-Methoden zur Verbesserung des diagnostischen Prozesses vorgetäuschter Symptome (Chafetz et al., 2015). Die diagnostischen Fragestellungen, mittels welcher Kriterien und Verfahren sich die BV verbessern lässt, sollten allerdings nicht ohne Berücksichtigung der Konsequenzen vorgetäuschter Symptome betrachtet werden. Neben finanziellen und medizinischen Konsequenzen ist auch der moralisch-zwischenmenschliche Aspekt zu berücksichtigen, so dass Täuschungsabsichten eine Gefährdung der therapeutischen Beziehung darstellen und die weitere Behandlung erschweren können. Demzufolge ist es Aufgabe des Behandlers, das Vorliegen von Täuschungsabsichten nicht automatisch mit dem Fehlen weiterer Pathologien gleichzusetzen, sondern entgegen der Reaktionen gegenüber dem Hirtenjungen bei Aesop die Täuschungsabsichten in den therapeutischen Prozess zu integrieren. Dabei kann der bedachte Einsatz von Methoden der BV einen relevanten Beitrag in der Behandlung psychischer und kognitiver Störungen leisten sowie Fehldiagnosen und ineffiziente Behandlungsstrategien begrenzen. Obwohl sich die diagnostischen und empirischen Fragestellungen der vorliegenden Arbeit mit spezifischen und detaillierten Methoden der neuropsychologischen BV befassen, sind sie als ein Bestandteil des klinischen Diagnoseprozesses zu betrachten

Das Ziel der vorliegenden Arbeit ist der Vergleich verschiedener diagnostischer Verfahren zur Erfassung verzerrter kognitiver Symptomdarstellungen in neuropsychologischen

Untersuchungen. In einer Analogstudie werden klinisch unauffällige Probanden zur Bearbeitung von Aufgaben unterschiedlicher kognitiver Funktionsbereiche (verbales und figürliches Gedächtnis, Aufmerksamkeit, Exekutivfunktionen, Visuokonstruktion) entweder mit maximaler Anstrengungsbereitschaft (Kontrollgruppe KG) oder zur Simulation kognitiver Beeinträchtigungen (experimentelle Simulanten ES) instruiert. Die kognitiven Leistungsprofile der beiden Gruppen werden zur Verbesserung der externalen Validität mit den Profilen von Patienten mit Simulationsverdacht verglichen. Dabei kommen neben speziellen Verfahren zur Erfassung der Anstrengungsbereitschaft (Performanzvalidierungstests, PVT) auch neuropsychologische Standardverfahren zum Einsatz, um aus sogenannten eingebetteten Parametern Hinweise über das Vorliegen verzerrter kognitiver Symptomdarstellungen abzuleiten. In einer weiteren Stichprobe mit kognitiv beeinträchtigten Patienten stehen die diagnostischen Grenzen der PVT im Fokus. Die kognitiven Beeinträchtigungen ermöglichen Rückschlüsse auf die hierfür beanspruchten kognitiven Funktionsbereiche und markieren die Grenzen der Anwendungsbereiche der PVT.

Die Studie leistet für mehrere, bisher nicht oder kaum untersuchte Aspekte der neuropsychologischen BV-Forschung relevante Beiträge. Das gängige experimentelle Untersuchungsverfahren anhand von Analogstudien mit Fallvignetten wird um die zweimalige Erinnerung an die Untersuchungsinstruktion ergänzt, da in Analogstudien nicht alle Probanden die Instruktion umsetzen (Bianchini, Mathias & Greve, 2001; Merten, 2014; Nies & Sweet, 1994). Darüber hinaus wird die diagnostische Güte des bislang wenig untersuchten *Aggravations- und Simulationstest (AST)*: Eberl & Wilhelm, 2007) ermittelt und sogenannte eingebettete Parameter aus Gedächtnis- und Aufmerksamkeits-tests analysiert, die bislang im deutschen Sprachraum wenig untersucht wurden. Durch die gezielte Analyse eingebetteter Parameter in kognitiven Standardverfahren soll ein ökonomischer Zugang zur BV für die klinische Alltagspraxis ermöglicht werden. Darüber hinaus werden die Grenzen der Aussagekraft der PVT und die zugrunde liegenden kognitiven Funktionen durch den Einbezug einer Stichprobe mit kognitiv beeinträchtigten Patienten spezifiziert.

2 Beschwerdenvalidierung in der Neuropsychologie

2.1 Begriffsbestimmung

In diesem Abschnitt werden die begrifflichen Differenzierungen der *Beschwerdenvalidierung* (BV) dargestellt. Unter BV versteht man die Überprüfung der Plausibilität und Konsistenz präsentierter Symptome und ermöglicht die Beurteilung der *Beschwerdenvalidität* (Merten, 2014). Da sowohl Symptomschilderungen als auch die Testperformanz hinsichtlich ihrer Beschwerdenvalidität überprüft werden, unterscheidet Larrabee (2012) zwischen *Symptomvalidität* (*symptom validity*) und *Performanzvalidität* (*performance validity*). Symptomvalidität verdeutlicht die Erfassung der Authentizität von Symptomschilderungen in Fragebögen, Performanzvalidität die Überprüfung der Glaubhaftigkeit präsentierter Testleistungen in neuropsychologischen Verfahren. Die Operationalisierung der BV erfolgt somit mittels diverser Methoden. Neben eingebetteten BV-Indikatoren, die aus neuropsychologischen Leistungstests abgeleitet werden, kommen auch speziell zur Erfassung der Performanzvalidität konstruierte Testverfahren, sogenannte *Performanzvaliditätstests* (PVT) zum Einsatz (Merten, 2014; Sweet & Breting, 2013). PVT sind neuropsychologische Testverfahren, die augenscheinlich kognitive Leistungstests darstellen, de facto aber die Anstrengungsbereitschaft erfassen (Merten, 2005). Ferner werden Selbstauskünfte in Fragebögen zur Symptomvalidierung herangezogen. Analog wird zwischen *Performanzvaliditätstests* (PVT) und *Selbstauskunfts-Validitätstest* (SRVT) unterschieden (Merten, Merckelbach, Giger & Stevens, 2016), die zuvor in der BV-Literatur als Beschwerdenvalidierungstests zusammengefasst wurden. In der vorliegenden Arbeit wird der Begriff BV, wie bei Merten et al. (2013), als übergeordneter Begriff verwendet.

Im Rahmen der BV werden *Antwortverzerrungen* (*response bias*) erfasst. Merten (2014) summiert darunter sämtliche Verhaltensweisen (Antworten, Selbstauskünfte und Testleistungen), die nicht den tatsächlichen Zustand einer Person widerspiegeln. Man differenziert zwischen *positiven* und *negativen Antwortverzerrungen*. Bei positiven Antwortverzerrungen (*positive response bias*) werden bestehende Beschwerden in abgeschwächter Form präsentiert. Diese Bagatellisierung wird häufig im Zusammenhang einer *Dissimulation* (*fake good*) beobachtet. Im Gegensatz dazu werden bei negativen Antwortverzerrungen (*negative response bias*) Symptome verstärkt dargestellt (Merten, 2014). Negative Antwortverzerrungen im neuropsychologischen Kontext werden auch als

eingeschränkte Anstrengungs- oder Leistungsbereitschaft (Merten, 2009), *suboptimale Leistungsmotivation* (Merten, 2009) oder *nicht-authentisches Testverhalten* (Bender & Rogers, 2004) bezeichnet, wobei die präsentierte Testperformanz unterhalb des tatsächlichen Leistungsniveaus liegt (Merten, Stevens & Blaskewitz, 2007). Heilbronner et al. (2009) begreifen Anstrengungsbereitschaft als Kontinuum (*effort level*), was sich im Verlauf einer Untersuchung, innerhalb und über verschiedene Testverfahren hinweg, verändern kann. Anstrengungsbereitschaft gilt als Voraussetzung zur Ermittlung valider Leistungsparameter in neuropsychologischen Untersuchungen und sollte idealerweise mehrmals im Verlauf überprüft werden (Heilbronner et al., 2009).

Eine weitere relevante Begriffsklärung ist die Unterscheidung zwischen BV und Simulationsdiagnostik, die im ersten Moment synonym erscheinen mögen, jedoch nicht gleichzusetzen sind. Im Gegensatz zur Simulationsdiagnostik erlaubt das Vorliegen verminderter Anstrengungsbereitschaft keine direkten, differentialdiagnostischen Rückschlüsse auf die zugrunde liegende Ätiologie oder Pathologie (Merten, 2011). Zwar können Neuropsychologen Simulationen diagnostizieren, dafür ist jedoch die Beurteilung weiterer Kriterien notwendig (Heilbronner et al., 2009; Merten & Dohrenbusch, 2010). Die Begriffe der BV werden in Tabelle 1 zusammengefasst und ergänzt.

Tabelle 1
Begriffe und Definitionen der BV.

Begriff	Definition
Aggravation	Übertriebene Darstellung tatsächlich bestehender Beschwerden
Anstrengungsbereitschaft (<i>effort</i>)	Leistungsmotivation, mit der Testaufgaben bearbeitet werden
Antwortverzerrungen	Manipulative Verhaltensweisen, die durch nicht-authentische Antworten gekennzeichnet sind und nicht das tatsächliche kognitive Leistungsniveau widerspiegeln (Testperformanz, Selbstauskünfte)
Beschwerdvalidität	Glaubhaftigkeit präsentierte Symptome
Beschwerdvalidierung (BV)	Prozess, der sämtliche Methoden der Plausibilitäts- und Konsistenzprüfung umfasst und in der Beschwerdvalidität resultiert
Dissimulation (<i>fake good</i>)	Bagatellisierung tatsächlich bestehender Beschwerden
Performanzvaliditätstest (PVT)	Speziell zur Erfassung negativer Antwortverzerrungen entwickelte Testverfahren (zuvor Beschwerdvalidierungstests genannt)
Simulation (<i>Malingering</i>)	Bewusste Symptomvortäuschung zur Erreichung eines externalen Zieles
Simulative Tendenzen	Aggravierende und simulative Verhaltensweisen zusammengefasst
Selbstauskunfts-Validitätstest (SRVT)	Standardisierte Erfassung von Selbstauskünften zur Symptompräsentation (Fragebögen, Interviews)

2.2 Antwortverzerrungen in diversen Kontexten

Antwortverzerrungen werden primär als Begleiterscheinungen psychischer Störungen in diversen Kontexten beschrieben und schon im Kindesalter beobachtet (Blaskewitz, Merten & Kathmann, 2008; Green & Flaro, 2003). Sie treten bei wahnhaften oder halluzinatorischen Störungsbildern, Persönlichkeitsstörungen und v.a. artifiziellen Störungen, Simulation, Aggravation sowie somatoformen und dissoziativen Störungen auf (Heubrock, Scholl & Petermann, 2013). Aktuell werden Antwortverzerrungen vermehrt auch in den Bereichen Intelligenzminderung und Aufmerksamkeitsdefizit-Hyperaktivitätsstörung untersucht (Hirsch & Christiansen, 2015). Neben psychischen und somatischen Beschwerden können auch kognitive Beeinträchtigungen Gegenstand negativer Antwortverzerrungen sein. Antwortverzerrungen können auch situationsbedingt durch mangelndes Interesse an der Testsituation, Müdigkeit, generelles Misstrauen gegenüber psychologischen Testverfahren, psychotrop wirkende Medikamente und Suchtmittel hervorgerufen werden (Stevens, Friedel, Mehren & Merten, 2008). Auch die Interaktionsdynamik zwischen Behandler und Patient kann die Symptompräsentation beeinflussen (Seron, 2014). Instruktionswidrige Verhaltensweisen und Antwortverzerrungen können u.a. als Reaktion auf einen desinteressierten oder wenig freundlichen Behandler resultieren (Widder, Dertwinkel, Egle, Foerster & Schiltewolf, 2007). Insgesamt ist von einer Vielzahl und möglichen Koexistenz mehrerer Einflussgrößen auszugehen. Heubrock et al. (2013) schlagen eine Unterscheidung zwischen authentischen und nicht-authentischen Störungen vor. Die wichtigsten Störungsgruppen nicht-authentischer Symptompräsentationen werden in Abbildung 1 verdeutlicht und im Folgenden dargestellt.

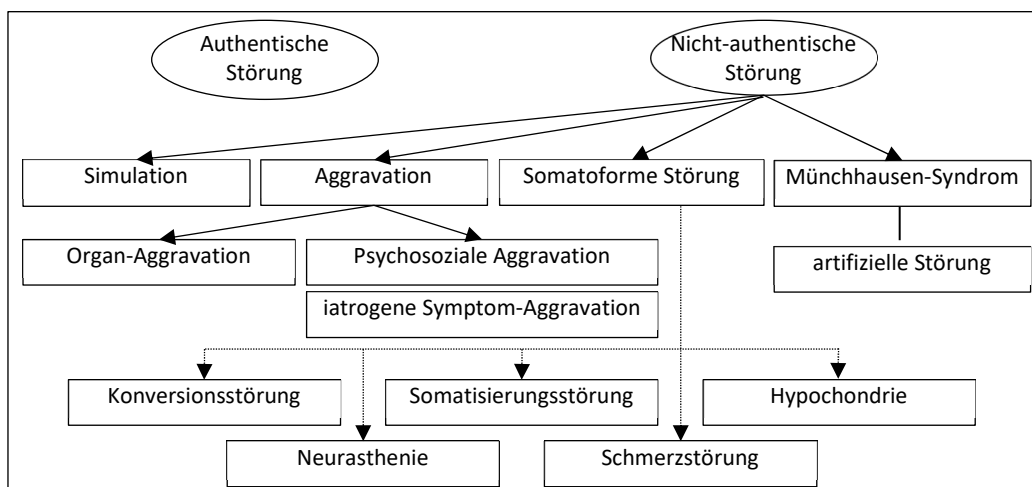


Abbildung 1. Differenzierung nicht-authentischer Störungen nach Heubrock et al. (2013).

2.2.1 Simulation und Aggravation

Bereits im 2. Jahrhundert vor Christus berichteten Rabbiner laut Nies und Sweet (1994) über Personen, die Beeinträchtigungen vortäuschten, um Unterstützung aus der Glaubensgemeinschaft zu erhalten (Stevens, 1986). Die Einführung des Simulationsbegriffs erfolgte nach Brussel und Hitch (1943) ursprünglich im militärischen Kontext zur Beschreibung von Soldaten, die ihre beruflichen Anforderungen krankheitsbedingt umgehen wollten (Nies & Sweet, 1994). *Simulation (Malingering)* wird auch als sekundärer Krankheitsgewinn bezeichnet (Merten, 2005). Dabei liegt der Vorteil nicht, wie beim primären Krankheitsgewinn, in der Krankenrolle oder Behandlung selbst, sondern außerhalb.

Die beiden Klassifikationssysteme *International Statistical Classification of Diseases and Related Health Problems (ICD-10: WHO, Dilling & Freyberger, 2014)* und *Diagnostic and Statistical Manual of Mental Disorders (DSM-5: APA, Falkai & Wittchen, 2014)* charakterisieren Simulation durch die absichtliche Produktion oder Präsentation körperlicher oder psychischer Beschwerden, motiviert durch externale Gründe wie Vermeidung von Strafverfolgung, Drogenbeschaffung, Vermeidung militärischer Verpflichtungen, monetäre Vorteile oder verbesserte Lebensbedingungen. Simulation wird allerdings nicht als eigenständiges Störungsbild definiert. Die ICD-10 kodiert Simulation unter „Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme von Gesundheitsdiensten führen“ als Z-Diagnose (Z76.5). Im DSM-5 wird Simulation im Abschnitt „Andere klinisch relevante Probleme“ aufgeführt, in dem Verhaltensweisen kategorisiert werden, die sich im Fokus der klinischen Aufmerksamkeit befinden oder die andere psychische Erkrankungen beeinflussen. Im Gegensatz zur ICD-10 werden im DSM-5 weitere Kriterien wie situationsbedingt adäquates Anpassungsverhalten (z.B. Geiselnahme), verminderte Kooperationsbereitschaft in Diagnostik und Behandlung (v.a. im forensischen Kontext), Diskrepanzen zwischen beschriebenen und objektivierbaren Defiziten und das Vorliegen einer Antisozialen Persönlichkeitsstörung erläutert. Zusammengefasst gelten ein objektivierbarer, externaler Anreiz und die Intentionalität der Symptomdarstellung als zentrale Kennzeichen einer Simulation (Merten, 2002, 2005).

Insgesamt sind die bestehenden Diagnosekriterien jedoch wenig spezifisch. So kann auf Grund überlappender Kriterien, wie der Diskrepanz zwischen subjektivem Krankheitserleben und objektivierbaren medizinischen Befunden, v.a. bei somatoformen und dissoziativen Störungsbildern fälschlicherweise der Verdacht einer Simulation entstehen (Merten, 2014). Zudem ist auch ein monetärer sekundärer Krankheitsgewinn, bedingt durch die

Absicherungen des Sozialstaates in der Bundesrepublik Deutschland, bspw. somatoformen oder artifiziellen Störungen zu finden (Merten, 2002).

Darüber hinaus erweist sich die differentialdiagnostische Abgrenzung zwischen Simulation und *Aggravation* als schwierig. Während bei einer Simulation Beschwerden vollständig vorgetäuscht werden, liegt bei einer Aggravation eine tatsächliche Erkrankung zugrunde, die verstärkt dargestellt wird. Die Schwierigkeit den authentischen Kern einer Simulation zu diagnostizieren und zwischen tatsächlichen Beeinträchtigungen und Simulation zu differenzieren, setzt die Kooperationsbereitschaft des Patienten voraus. Allerdings ist anzunehmen, dass ein Simulant seine vorgetäuschten Symptome glaubhaft aufrechterhalten möchte und keine zuverlässige Unterscheidung zwischen Aggravation und Simulation ermöglicht (Merten, 2014). Aus diesem Grund schlägt Wurzer (1992) den Zusammenschluss der beiden Phänomene zu *simulativen Tendenzen* vor. Ferner liegen sowohl der Simulation als auch der Aggravation Täuschungsabsichten (Motive) zugrunde, die das entsprechende Verhalten hervorrufen (Heubrock et al., 2013; Merten, 2009).

Aggravation wird weder in der ICD-10 noch im DSM-5 als eigenständiges Störungsbild definiert. Eine Differenzierung des Aggravationsbegriffs anhand bewusster und unbewusster Motive schlagen dagegen Heubrock et al. (2013) vor. Als *Organ-Aggravation* bezeichnen die Autoren eine extern motivierte Aggravation. Zum Erhalt ungerechtfertigter, objektivierbarer Vergünstigungen werden bestehende Symptome verstärkt dargestellt. Bei der *psychosozial motivierten Aggravation* erfolgt die Beschwerdenübertreibung, um auf die Erkrankung und deren Auswirkungen aufmerksam zu machen. Speziell im Hinblick auf jahrelange gerichtliche Auseinandersetzungen führen Heubrock et al. (2013) den Begriff der *iatrogenen Symptom-Aggravation* ein, der die aggravationsfördernden Interaktionsprozesse zwischen Diagnostiker und Patient bezeichnet. Beim Patienten können durch diverse Untersuchungen Zweifel entstehen, von Behandlern, Gutachtern und gerichtlichen Institutionen ernst genommen zu werden. Dies kann in der Übertreibung bestehender Beeinträchtigungen in weiteren Untersuchungen resultieren (Heubrock et al., 2013). Aggravationstendenzen gelten gemäß Expertenschätzungen als Bestandteil des klinischen Alltags, die Häufigkeit wird bspw. im Schweizer Rehabilitationskontext auf 10% geschätzt (Frei, 2004, zitiert nach Merten, Friedel & Stevens, 2006).

2.2.2 Artificielle Störung

Die *artificielle Störung* wird definiert als das absichtliche Erzeugen oder Vortäuschen körperlicher oder psychischer Symptome oder Behinderungen ohne erkennbaren externalen

Anreiz (Dilling & Freyberger, 2014). Die zugrunde liegende Motivation für die zum Teil stark selbstschädigenden Verhaltensweisen (Unterzuckerung, Wundmanipulation, Fiebererzeugung, Manipulation des Enddarms oder der Haut) ist meist unklar. Es ist allerdings anzunehmen, dass das Einnehmen einer Krankenrolle und des damit verbundenen Krankheitsgewinns intendiert wird. Durch die aktive Symptomproduktion des Patienten verlässt er seine passive Opferrolle und behindert gleichzeitig, ohne Wissen des Behandlers, eine erfolgreiche Therapie (Eckhart, 1996). Die Störung weist eine hohe Komorbidität mit Persönlichkeits- und Beziehungsstörungen auf.

In der ICD-10 wird die artifizielle Störung (F68.1) den *Persönlichkeits- und Verhaltensstörungen* (F68) zugeordnet, ebenso die *Rentenneurose* synonym der *Entwicklung körperlicher Symptome aus psychischen Gründen* (F68.0). Dazu zählen körperliche Symptome mit Hinweisen einer psychischen Ätiologie, die in Folge einer Erkrankung oder Behinderung aggraviert werden. Zu den artifiziellen Störungen zählen nach ICD-10 neben *durch Institutionen wandernde Patienten* auch *Krankenhauspringer* (*hospital hopper syndrome*) und das *Münchhausen-Syndrom*. Das Münchhausen-Syndrom ist durch die pseudologische Ausgestaltung einer umfangreichen Krankheitsgeschichte, soziale Entwurzelung und exzessives Wandern von Klinik zu Klinik gekennzeichnet. Die *artifizielle Störung in Vertretung*, auch *artifizielle Störung by proxy* oder *erweitertes Münchhausen-Syndrom* genannt, wird in der ICD-10 nicht den artifiziellen Störungen, sondern der Kindesmisshandlung (T74.8) zugeordnet (Dilling & Freyberger, 2014). Im DSM-5 wird die artifizielle Störung als *vorgetäuschte Störung* unter den *Somatischen Belastungsstörungen und verwandten Störungen* aufgeführt. Die Prävalenzschätzungen vorgetäuschter Verhaltensweisen liegen im stationären Kontext bei ca. 1% (Falkai & Wittchen, 2014).

2.2.3 Somatoforme und dissoziative Störungen

Das allgemeine Kennzeichen *somatoformer Störungen* sind körperliche Beschwerden mit beharrlichen Behandlungsforderungen, die nicht vollständig durch einen medizinischen Krankheitsfaktor erklärt werden können (Sauer & Eich, 2007). Somatoforme Störungen (F45) sind in der ICD-10 u.a. ein Überbegriff für die Störungsbilder *Somatisierungsstörung*, *undifferenzierte Somatisierungsstörung*, *somatoforme autonome Funktionsstörung*, *hypochondrische Störung* (einschließlich *körperdysmorphe Störung*), *Schmerzstörung*, *Fibromyalgie* (Schmerzen in unterschiedlichen Körperregionen), *Neurasthenie* (anhaltende übersteigerte Erschöpfung) und *chronisches Müdigkeitssyndrom* (*chronic fatigue syndrome*) (Deister, 2013; Dilling & Freyberger, 2014). Den *Dissoziativen Störungen* (F44) liegt ein

Integrationsverlust von Erinnerungen der eigenen Identität, unmittelbarer Empfindungen oder Körperbewegungen zu Grunde (Deister, 2013). Dazu zählen die *dissoziative Amnesie*, *dissoziative Fugue*, *dissoziative Störung der Bewegung und der Sinnesempfindung (Konversionsstörung)* sowie *andere dissoziative Störungen (Ganser-Syndrom, multiple Persönlichkeitsstörung)*. Der *dissoziative Stupor* (Unfähigkeit zu Willkürbewegungen und Reaktionen auf äußere Reize) und *Trance- und Besessenheitszustände* werden in der ICD-10 ebenfalls dieser Störungsgruppe zugeordnet (Dilling & Freyberger, 2014). Im DSM-5 werden somatoforme Störungen als *Somatische Belastungsstörung und verwandte Störungen* bezeichnet, denen die Diagnosen der *Psychologischen Faktoren, die eine medizinische Krankheit beeinflussen* und auch die *vorgetäuschte Störung* und die *Konversionsstörung* zugeordnet werden. Die Lebenszeit-Prävalenzen *Somatischer Belastungsstörungen* liegen bei 5 bis 7%, wahrscheinlich mit höherem Frauenanteil. Bei ca. 5% der Patienten in der sekundären neurologischen Krankheitsversorgung werden Konversionsstörungen diagnostiziert (Falkai & Wittchen, 2014).

2.2.4 Differentialdiagnostische Unterscheidung

Zur Vereinfachung der Unterscheidung von Simulation / Aggravation, artifizieller Störung und somatoformer bzw. dissoziativer Störung führen Slick et al. (1999) die Differenzierung mit zwei dichotomen Bewertungsdimensionen ein (Tabelle 2). Während die erste Dimension die Bewusstseinsnähe bzw. Intentionalität (bewusst intendiert vs. unbewusst) der Symptomproduktion darstellt, bildet die zweite Dimension die zugrunde liegende Motivation (external vs. internal psychologisch) ab. Bei einer Simulation sind sowohl die Symptomproduktion als auch die zugrundeliegende Motivation der Person bewusst. Im Rahmen einer artifiziellen Störung ist die Symptomproduktion der Person bewusst, die zugrundeliegende Motivation ist dagegen unbewusst und unreflektiert. Bei somatoformen und dissoziativen Störungen sind weder die Symptomproduktion noch die zugrundeliegende Motivation der Person bewusst (Merten, 2011).

Tabelle 2

Klassifikation nicht-authentischer Störungen anhand Bewusstseinsgrad und Motivation (modifiziert nach Merten (2011) und Slick et al. (1999)).

Diagnosegruppe	Symptomproduktion	Motivation	Beispiele
Simulation	bewusst: intendiert, reflektiert, willentlich	bewusst: reflektiert, sekundärer Krankheitsgewinn	Finanzielle Entschädigung, Vermeidung von Sanktionen oder Verpflichtungen
Artifizielle Störung	bewusst: intendiert, reflektiert, willentlich	unbewusst: unreflektiert, primärer Krankheitsgewinn	Krankenrolle, Zuwendung, nicht-monetäre Verstärkung, Vermeidung, Stressreduktion
Somatoforme/ dissoziative Störung	unbewusst: nicht intendiert, nicht willentlich	unbewusst: unreflektiert	Lösungsversuch für Stress- oder Konfliktsituation

Trotz der Eindeutigkeit der Kriterien in der Theorie, gestaltet sich die Klassifikation in der Praxis häufig als schwierig (Slick et al., 1999). Die Zuordnung basiert auf der Beurteilung der innerpsychischen Zustände der Person, die dem Behandler nicht direkt zugänglich sind (Merten, 2011) und ist mit einer diagnostischen Unsicherheit verbunden. Demzufolge erscheint die Erweiterung des diagnostischen Prozesses durch standardisierte Verfahren als hilfreich.

Im Rahmen der BV werden Antwortverzerrungen zur Beurteilung der Performanzvalidität mittels kognitiven Leistungstests und der Symptomvalidität mittels Selbstauskünften erfasst. Antwortverzerrungen können diverse Faktoren zugrunde liegen und treten in verschiedenen Kontexten auf. Nicht-authentische Symptompräsentationen werden u.a. bei artifiziellen Störungen, Simulation, Aggravation sowie somatoformen und dissoziativen Störungen beobachtet. Im Hinblick auf die differentialdiagnostischen Schwierigkeiten erweist sich eine Klassifikation nach Intentionalität und des Motivs der Symptomproduktion als hilfreich. Insgesamt ist die Überprüfung der Beschwerdvalidität diagnostisch und folglich auch empirisch als schwer operationalisierbarer Aspekt der klinischen Praxis zu betrachten.

2.3 Diagnostische Kriterien der Simulation kognitiver Beeinträchtigungen

Lange galt als einzig sicheres Kriterium für das Vorliegen einer Simulation das entsprechende Eingeständnis der Person. Allerdings wird ein Simulant versuchen, sein Täuschungsverhalten zur Sicherung der erhofften Vorteile aufrechtzuerhalten und eine Diagnosestellung muss unabhängig eines Eingeständnisses erfolgen (Merten, 2014). Auf Grund mangelnder Spezifität

und unzureichendem Forschungsbezug der ICD-10 und des DSM erfolgten in den 1990er Jahren diverse Erweiterungen und Spezifizierungen durch neuropsychologische und verhaltensbezogene Kriterien u.a. von Faust und Ackley (1998) oder Greiffenstein, Baker und Gola (1994). Rogers (1990b) führte die explizite Unterscheidung zwischen der Simulation psychischer Störungen, körperlicher Erkrankungen und kognitiver Beeinträchtigungen ein.

2.3.1 Diagnostische Kriterien nach Slick

Die wichtigsten diagnostischen Richtlinien zur Erfassung vorgetäuschter kognitiver Störungen stellen die Kriterien von Slick et al. (1999) dar. Sie liefern einen maßgeblichen konzeptionellen Beitrag für die Diagnosestellung kognitiver Simulationen und liegen zahlreichen Untersuchungen zu Grunde (Bianchini, Curtis & Greve, 2006; Heilbronner et al., 2009; Merten, 2011; Stevens et al., 2008). Zur Erfassung simulativer Verhaltensweisen definierten Slick et al. (1999) die *Simulation neurokognitiver Dysfunktionen* (SND) als bewusstes Übertreiben oder Vortäuschen kognitiver Beeinträchtigungen, um materielle Vergünstigungen zu erhalten oder formale Pflichten und Verantwortung zu vermeiden. Die Diagnostik stützt sich auf vier Kriterien:

1. externaler Anreiz (Kriterium A)
2. auffällige neuropsychologische Testleistung (Kriterium B)
3. inkonsistente Beschwerdenanamnese (Kriterium C)
4. nicht gänzlich anders erklärbare Verhaltensweisen (Kriterium D)

Slick et al. (1999) berücksichtigten neben der Intentionalität der Symptomproduktion auch die diagnostische Sicherheit, analog zur Demenzdiagnostik (McKhann et al., 1984), durch die Einteilung in sicher, wahrscheinlich und möglich. Obwohl es sich dabei um intendierte Verhaltensweisen handelt, erlaubt deren Erfassung jedoch keine direkten Rückschlüsse auf die zugrunde liegende Absicht. Im Folgenden werden die Slick-Kriterien in Anlehnung an die deutschen Übersetzungen von Eberl und Wilhelm (2007) und Merten (2014) in Tabelle 3 zusammengefasst und beschrieben.

Tabelle 3

Beschreibung der Kriterien nach Slick et al. (1999).

Kriterium A: Vorliegen eines substanziellen externalen Anreizes

Mindestens ein klar identifizierbarer und außerhalb der Störung liegender Gewinn muss zum Zeitpunkt der Untersuchung vorhanden sein.

Kriterium B: Neuropsychologische Untersuchungsergebnisse deuten auf Antwortverzerrungen

B1. Testleistung unterhalb der Ratewahrscheinlichkeit in mindestens einem Alternativwahlverfahren

B2. Auffällige Testleistung in mindestens einem gut validierten psychometrischen Verfahren der BV (Test oder Indikator)

B3. Diskrepanzen zwischen den Testleistungen und bekannten zerebralen Funktionsmustern: Das neuropsychologische Leistungsprofil weicht von aktuell anerkannten Modellen normaler und abnormaler kognitiver Funktionsweisen ab

B4. Diskrepanzen zwischen den Testleistungen und der Verhaltensbeobachtung: Die Ergebnisse in mind. einem neuropsychologischen Testverfahren sind nicht mit dem beobachteten kognitiven Funktionsniveau vereinbar

B5. Diskrepanzen zwischen den Testleistungen und zuverlässigen fremdanamnestischen Informationen: Die Ergebnisse in mind. zwei neuropsychologischen Testverfahren sind nicht mit den Angaben über das alltagsbezogene Funktionsniveau des Patienten durch mind. einen Dritten vereinbar

B6. Diskrepanzen zwischen den Testleistungen und anamnestischen Angaben in der Patientenakte: Die Ergebnisse in mind. zwei neuropsychologischen Testverfahren eines kognitiven Funktionsbereichs sind nicht mit der neurologischen oder psychiatrischen Vorgeschichte des Patienten vereinbar

Kriterium C: Beschwerdenschilderungen deuten auf Antwortverzerrungen

C1. Diskrepanzen zwischen den Schilderungen durch den Patienten und der dokumentierten medizinischen oder psychosozialen Vorgeschichte in der Patientenakte

C2. Diskrepanzen zwischen den Beschwerdenschilderungen und bekannten zerebralen Funktionsmustern: Die berichteten Symptome sind nicht mit Art und Schweregrad der Verletzung oder Psychopathologie vereinbar

C3. Diskrepanzen zwischen den Beschwerdenschilderungen und der Verhaltensbeobachtung

C4. Diskrepanzen zwischen den Angaben des Patienten und zuverlässigen fremdanamnestischen Informationen: Die berichteten Symptome, die Vorgeschichte oder das gezeigte Verhalten sind nicht mit Angaben zuverlässiger Dritten vereinbar

C5. Evidenz für übertriebene oder erfundene psychologische Defizite: fragwürdige Symptomvalidität in gut validierten Selbstbeurteilungsverfahren (Fragebögen, standardisierte Interviews)

Kriterium D: Ausschluss anderer Faktoren

Verhaltensweisen, die sich den Kriterien B und C zuordnen lassen, allerdings nicht vollständig durch eine psychiatrische, neurologische oder entwicklungsbedingte Störung erklären lassen.

2.3.1.1 Sichere Simulation neurokognitiver Dysfunktionen (SND)

Eine *sichere SND* liegt bei eindeutigen und zwingenden Hinweisen für bewusstes Übertreiben oder Vortäuschen kognitiver Defizite bei fehlenden plausiblen Alternativerklärungen vor. Die obligatorischen und spezifischen Kriterien einer sicheren SND sind:

- Vorhandensein eines substanziellen externalen Anreizes (Kriterium A)
- Eindeutiges Vorliegen negativer Antwortverzerrungen in einem neuropsychologischen Testverfahren, d.h. die Testleistung in einem oder mehreren PVT liegt unterhalb der Ratewahrscheinlichkeit ($p < .05$) (Kriterium B1)
- Die Verhaltensweisen (Kriterium B) sind nicht vollständig durch psychiatrische, neurologische oder entwicklungsbedingte Faktoren erklärbar (Kriterium D)

In der neuropsychologischen Forschungsliteratur hat sich ein Konsens für die Definition einer sicheren SND gebildet, der auch von der DGPPN geteilt wird (Dressing et al., 2011). Eine Testleistung unterhalb des Zufallsniveaus ($p < .05$) wird als sichere Evidenz für ein bewusstes Übertreiben kognitiver Defizite betrachtet, da dies als aktives Vermeiden der richtigen Antwort gilt (Bianchini, Greve & Glynn, 2005; Dressing et al., 2011; Larrabee, Greiffenstein, Greve & Bianchini, 2007). In der Praxis sind solche Ergebnisse jedoch selten. In einer Untersuchung von Gervais, Rohling, Green und Ford (2004) zeigten beispielsweise weniger als 2% der 519 Patienten mit Simulationsverdacht eine Testleistung unterhalb des Zufallsniveaus. Als weiteres sicheres Kriterium einer SND gilt eine *überwältigende Inkonsistenz (compelling inconsistency)* aus der Beobachtung eines Verhaltens, das zuvor als ausgeschlossen beschrieben wurde (Schmidt et al., 2011). Beispielhaft sind Einschränkungen der Arbeitsfähigkeit in einer klinischen Untersuchung, bei gleichzeitigem Vorliegen eindeutiger Beweise der Arbeitsfähigkeit wie Zeugenaussagen, Videos, etc. (Bianchini et al., 2005).

2.3.1.2 Wahrscheinliche Simulation neurokognitiver Dysfunktionen

Eine *wahrscheinliche SND* ist durch einen starken Hinweis für das bewusste Übertreiben oder Vortäuschen kognitiver Defizite bei fehlenden plausiblen Alternativerklärungen gekennzeichnet. Die obligatorischen und spezifischen Kriterien sind durch zwei mögliche Kriterienkonstellationen definiert:

- Vorliegen eines substanziellen externalen Anreizes (Kriterium A)
- Mindestens zwei Indizien für negative Antwortverzerrungen aus neuropsychologischen Verfahren (Kriterium B2 - B6)
oder
ein Kriterium aus B2 - B6 und mindestens eine negative Antwortverzerrung in Selbstauskünften (Kriterium C1 - C5)

- Die Verhaltensweisen (Kriterium B und C) sind nicht vollständig durch psychiatrische, neurologische oder entwicklungsbedingte Faktoren erklärbar (Kriterium D)

Laut Larrabee et al. (2007) besteht in der Literatur auch ein Konsens hinsichtlich der Diagnosevergabe einer wahrscheinlichen SND.

2.3.1.3 Mögliche Simulation neurokognitiver Dysfunktionen

Eine *mögliche SND* zeichnet sich durch Hinweise für das bewusste Übertreiben oder Vortäuschen kognitiver Defizite bei fehlenden plausiblen Alternativerklärungen aus. Eine mögliche SND ist anzunehmen, wenn die Kriterien einer sicheren oder wahrscheinlichen Simulation erfüllt sind, allerdings mit der Einschränkung, dass andere primäre Ursachen nicht ausgeschlossen werden können. Die obligatorischen und spezifischen Kriterien einer wahrscheinlichen SND sind durch zwei mögliche Kriterienkonstellationen definiert:

- Vorliegen der Kriterien einer sicheren oder wahrscheinlichen SND mit Ausnahme von Kriterium D, so dass alternative Ursachen für die Antwortverzerrungen angenommen werden, die weiter spezifiziert werden müssen
oder
- Vorliegen eines substantiellen externalen Anreizes (Kriterium A) und mindestens eine Antwortverzerrung in den Selbstauskünften (Kriterium C1 - C5) und die Verhaltensweisen (Kriterium C) sind nicht vollständig durch psychiatrische, neurologische oder entwicklungsbedingte Faktoren erklärbar (Kriterium D)

2.3.2 Modifikation und Erweiterung der Kriterien nach Slick

Eine kritische Reflexion der Slick-Kriterien findet sich bei Larrabee et al. (2007). Als Vorteile nennen die Autoren u.a. den multimodalen Zugang, die systematische und strukturierte Anwendung, die Adaptationsmöglichkeiten für verschiedene Formen kognitiver Simulation und das konservative Vorgehen zur Verringerung von Fehlklassifikationen. Als Nachteile werden der Fokus auf kognitive Simulationen, die Vernachlässigung anderer Verhaltensbereiche und die erhöhte Anfälligkeit für Verzerrungen durch den Kliniker genannt. Neben der Entwicklung klarer, objektiver Standards schlagen die Autoren auch eine Gleichsetzung der B- und C-Kriterien zur Steigerung der Sensitivität vor. Neben den inhaltlichen Entwicklungen zur Definition eindeutiger Diagnosekriterien wird das diagnostische Vorgehen auch aus einer konzeptionellen Perspektive betrachtet. Diverse

Autoren nehmen an, dass die Simulation kognitiver Defizite als Kontinuum zu betrachten ist (Bigler, 2015; Heilbronner et al., 2009; Walters, Berry, Rogers, Payne & Granacher, 2009). Simulation und Anstrengungsbereitschaft werden als dimensionale (kontinuierliche) statt taxonomische (kategoriale) Konstrukte begriffen, die durch eine kategorische Klassifikation nicht adäquat abgebildet werden. Bigler (2015) betont bei der Verwendung dichotomer Kriterien die Gefahr von Falsch-Positiv-Diagnosen. Auch Heilbronner et al. (2009) verstehen die Entwicklung diagnostischer Kriterien als deskriptive Begriffe zur Beschreibung bewusster Symptomverstärkungen und distanzieren sich von dem Anspruch einer vollständigen Erfassung des Krankheitsprozesses.

Seit den 1990er Jahren rückte die Entwicklung von Kriterien zur Simulationsdiagnostik in neuropsychologischen Untersuchungen in den Forschungsfokus der BV. Einen wesentlichen Beitrag liefern bis heute die Slick-Kriterien. Neben einem externalen Anreiz bilden die Kriterien verhaltensbezogene und neuropsychologische Diskrepanzen ab, die indirekt eine Intentionalität des Verhaltens andeuten. Die Einteilung dieser festgelegten Kriterien erfolgt unter Berücksichtigung der diagnostischen Sicherheit. Über die Modifikation bestehender Kriterien hinaus wird aktuell die klassifikatorisch-kategoriale um eine kontinuierliche Betrachtungsweise ergänzt.

2.4 Studiendesigns der BV

Die Forschungsarbeiten zur neuropsychologischen BV basieren auf drei Studiendesigns: *Analogstudien*, *known-group Studien* und *Prävalenzstudien* (Heilbronner et al., 2009). Bislang wurden die meisten Forschungsarbeiten als Analogstudien konzipiert, die als vergleichsweise einfach operationalisierbar gelten (Hilsabeck, LeCompte, Marks & Grafman, 2001). In jüngerer Zeit ist eine wachsende Anzahl an Untersuchungen mit *known-group Design* und Patienten mit Simulationsverdacht zu verzeichnen.

In Analogstudien werden klinisch unauffällige Personen randomisiert unterschiedlichen experimentellen Bedingungen (Kontrollgruppe vs. Experimentalgruppe) zugeordnet. In einigen Analogstudien werden die gesunden Probanden mit klinischen Stichproben, meist Patienten mit Schädelhirntrauma, verglichen (Willison & Tombaugh, 2006). Die Probanden der Experimentalgruppe werden zur Simulation kognitiver Beeinträchtigungen instruiert (*experimentelle Simulanten, ES*), die Patienten- und Kontrollgruppen werden zur authentischen Aufgabenbearbeitung aufgefordert. Die experimentellen Simulanten erhalten

zur Förderung des Instruktionsverständnisses meist eine Fallvignette eines potentiellen Simulanten (Merten, Henry & Hilsabeck, 2004). Auf Grund der hohen experimentellen Kontrolle und der standardisierten Vorgehensweise stellen Analogstudien ein effektive Methode zur konzeptionellen Validierung neuer Testverfahren dar (Heilbronner et al., 2009). Insgesamt liegt der Vorteil von Analogstudien in der hohen internen Validität begründet. Die externe Validität und die Generalisierbarkeit der Testleistungen der experimentellen Simulanten auf Patienten mit negativen Antwortverzerrungen ist dagegen eingeschränkt (Demakis, 2004; Heilbronner et al., 2009; Hilsabeck et al., 2001; Nies & Sweet, 1994). Laut Demakis (2004) liegt dies in der Zusammensetzung der Stichprobe aus gesunden Probanden und der Instruktionsumsetzung begründet. Es ist davon auszugehen, dass sich experimentelle und klinische Simulanten hinsichtlich demographischer Variablen (Alter, Geschlecht und Bildungsabschluss), der Angst entdeckt zu werden und der zugrundeliegenden Motivation unterscheiden. Zudem sind die externalen Anreize, wie Kompensationszahlungen, nicht mit den experimentellen Versuchungsanordnungen der Studienteilnahme vergleichbar. Darüber hinaus wird meist ausschließlich die Testperformanz untersucht, weitere klinisch relevante Informationen aus der Verhaltensbeobachtung werden vernachlässigt. Außerdem können klinische Simulanten, im Vergleich zu den klinisch unauffälligen Probanden, tatsächlich kognitive Beeinträchtigungen aufweisen, auch wenn die aggravierte Symptomdarstellung den tatsächlichen Schweregrad übersteigt (Demakis, 2004).

Im *known-group Design* werden Patienten anhand von a priori definierten Kriterien für negative Antwortverzerrungen (neuropsychologische Parameter, Persönlichkeitsmerkmale, Krankheitsgeschichte) selektiert (Demakis, 2004). Beispielsweise werden Patienten mit leichten Schädelhirnverletzungen ausgewählt, die in einem Gerichtsverfahren Kompensationszahlungen geltend machen und in einem PVT auffällige Werte erzielen (Heilbronner et al., 2009). Anschließend werden die Symptomdarstellungen dieser Gruppe mit einer Patientenstichprobe mit objektivierbaren Symptomen, meist erworbenen Schädel-Hirn-Verletzungen, verglichen. Der klinische Praxisbezug durch Patienten mit externalen Anreizen und negativen Antwortverzerrungen resultiert in einer hohen externen Validität des Designs (Hilsabeck et al., 2001). Dagegen ist auf Grund unzureichender standardisierter Experimentalbedingungen von einer geringen internen Validität auszugehen. Auch Fehlklassifikationen können nicht ausgeschlossen werden, da die wenigsten Patienten ihre Simulationsabsichten zugeben. Die Herausforderung besteht hier also in der Definition der Auswahlkriterien (Heilbronner et al., 2009).

Während sich die Vor- und Nachteile von Analogstudien und *known-group Studien* gegenseitig aufheben, können anhand von Prävalenzstudien weitere Informationen hinsichtlich der Häufigkeiten negativer Antwortverzerrungen ermittelt werden (Heilbronner et al., 2009). Da die Häufigkeit in Abhängigkeit des Kontextes und des Anreizes variiert, erfolgen in Prävalenzstudien die Selektion und der Vergleich der Stichproben anhand des Kontexts. Auf Grund der fehlenden Standardisierung solcher Untersuchungen ist deren interne Validität als niedrig zu bewerten. Eine weitere Einschränkung liegt in der unterschätzten Häufigkeit von Simulation durch die Verwendung konservativer Kriterien (Dandachi-FitzGerald, Ponds, Peters & Merckelbach, 2011).

2.5 Prävalenzraten

Die diagnostischen Schwierigkeiten bei der Erfassung bewusster und zielgerichteter Simulationen spiegeln sich auch in variierenden Prävalenzraten wider, die laut Dandachi-FitzGerald et al. (2011) zwischen 20% und 60% rangieren. Neben der diagnostischen Sicherheit beeinflussen auch Grunderkrankungen, der Kontext sowie die verwendeten Diagnosekriterien die Prävalenzschätzungen. Zur Ermittlung von Basisraten werden Prävalenzstudien und Expertenbefragungen herangezogen.

2.5.1 Kontextspezifische Prävalenzen

Begutachtungen sind situationsbedingt mit substantiellen externalen Anreizen wie finanziellen Entschädigungen oder strafrechtlichen Konsequenzen verbunden. Folglich ist mit höheren Prävalenzraten als in anderen Kontexten zu rechnen (Merten, 2014; Merten, Krahl, Krahl & Freytag, 2010). Die Prävalenzschätzungen simulativer Verhaltensweisen im deutschsprachigen Gutachtenkontext rangierten in einer Expertenbefragung von Dandachi-FitzGerald, Ponds und Merten (2013) zwischen 0 und 100%. Dabei zeigte sich eine Diskrepanz zwischen den Einschätzungen der eigenen Berufspraxis (Median = 15%) und der allgemeinen Berufspraxis (Median = 20%). Die Annahme, dass simulatives Verhalten im Rahmen von Begutachtungen niemals aufträte, vertraten hinsichtlich der eigenen Praxis 26% und im Allgemeinen 0.5% der Befragten (Dandachi-FitzGerald et al., 2015). Eine höhere Prävalenzrate von 48% ergab die retrospektive Datenanalyse von knapp 400 deutschsprachigen Gutachtenpatienten der gesetzlichen Unfallversicherung (Merten et al., 2010). In einer weiteren deutschsprachigen Prävalenzstudie mit 233 Gutachtenpatienten wurde eine vergleichbare SND-Basisrate von 44.6%, anhand von zwei auffälligen

unabhängigen PVT, ermittelt (Stevens et al., 2008). Die Prävalenzschätzungen sicherer SND im nordamerikanischen Gutachtenkontext lagen in einer Umfrage von Slick, Tan, Strauss und Hultsch (2004) zwischen 5% und 30% (Modus = 10% - 20%). Die Prävalenzschätzungen einer möglichen SND rangierten ebenfalls zwischen 5% und 30% (Modus = 5% - 10%).

Die wohl bekannteste Expertenumfrage wurde von Mittenberg, Patton, Canyock und Condit (2002) mit 131 praktisch tätigen nordamerikanischen Neuropsychologen durchgeführt. Die vielfach zitierten Prävalenzschätzungen dieser Studie resultierten aus 33 000 Gutachtenpatienten. Die höchsten Prävalenzen von 32.7% wurden bei Begutachtungen von Frühberentung oder Arbeitsunfähigkeit ermittelt. Bei zivilrechtlichen Begutachtungen lagen die Basisraten bei 30.4% und bei strafrechtlichen Begutachtungen bei 22.8%. Die niedrigsten Prävalenzen von 8.1% waren in medizinisch/ psychiatrischen Behandlungen zu verzeichnen. Neben dem Gutachtenkontext stellt auch die soziale Schichtzugehörigkeit einen relevanten Einflussfaktor dar. In einer Prävalenzstudie von Göbber, Petermann, Piegza und Kobelt (2012) mit 457 deutschstämmigen und 138 ausländischen Gutachtenpatienten der Deutschen Rentenversicherung mit medizinisch-psychosomatischen Beschwerden zeigten 33% der deutschen Versicherten und 50% der Versicherten mit Migrationshintergrund negative Antwortverzerrungen in einem SRVT. Die Gruppenunterschiede konnten nicht ausreichend durch den Migrationsstatus erklärt werden, sondern waren größtenteils auf Geschlechts- und Schichtzugehörigkeitsdifferenzen zurückzuführen.

Auch außerhalb des Gutachtenkontextes ist mit dem Auftreten nicht-authentischer kognitiver Symptompräsentationen zu rechnen. Laut einer Expertenumfrage von Sharland und Gfeller (2007) liegen die Prävalenzraten im klinischen Kontext bei 5%. Dandachi-FitzGerald et al. (2013) ermittelten in einer Umfrage mit neuropsychologischen Sachverständigen aus sechs europäischen Staaten (Deutschland, Italien, Dänemark, Finnland, Norwegen und Niederlande) Prävalenzschätzungen für die Simulation kognitiver Dysfunktionen im klinischen Kontext von 10%, in der eigenen Praxis von 4%. Die deutschsprachigen Teilnehmer ($n = 198$) schätzten die Prävalenzraten im klinischen Kontext auf 5% (Spannweite = 0 - 70%) (Dandachi-FitzGerald et al., 2015). In einer Prävalenzstudie mit einer psychiatrischen Stichprobe ohne externalen Anreiz ermittelten Dandachi-FitzGerald et al. (2011) eine vergleichbare Prävalenzrate von 8% anhand auffälliger Werte in einem PVT und einem SRVT. Kemp et al. (2008) erzielten eine etwas höhere Basisrate von 11% mit neurologischen Patienten, die medizinisch nicht erklärable Symptome und mindestens zwei auffällige BV-Indikatoren zeigten (Dandachi-FitzGerald et al., 2013). In einer Studie von Schroeder und Marshall (2011)

präsentierten 7% der Patienten mit psychotischer Symptomatik und 5% der psychiatrischen Patienten ohne psychotische Symptomatik in mindestens zwei PVT auffällige Ergebnisse oder in einem PVT ein Ergebnis unterhalb der Ratewahrscheinlichkeit. Obwohl im klinischen Kontext von niedrigeren Prävalenzen als im Begutachtungskontext auszugehen ist, scheinen verzerrte Symptomdarstellungen zur klinischen Routine zu gehören.

2.5.2 Diagnosespezifische Prävalenzen

Darüber hinaus unterscheiden sich einzelne Diagnosegruppen in der Häufigkeit und dem Ausprägungsgrad nicht-authentischer Symptompräsentationen. Bei Krankheitsbildern mit wenig spezifischen (kognitiven) Symptomen werden häufiger verzerrte Symptomdarstellungen beobachtet als bei schweren Störungen (Dohrenbusch, Henningsen & Merten, 2011). So ist bei posttraumatischen Belastungsstörungen (PTBS), Schmerzstörungen, Fibromyalgie, chronischen Erschöpfungssyndromen und umwelt-assoziierten Erkrankungen mit höheren Prävalenzraten zu rechnen. Am häufigsten werden negative Antwortverzerrungen bei Schädelhirntraumata berichtet (Merten et al., 2006). Larrabee (2003) ermittelte aus 11 Studien mit 1363 Schädelhirntrauma-Patienten eine mittlere Prävalenzrate von 40%. Die Prävalenzraten rangierten zwischen 15% (Trueblood & Schmidt, 1993) und 64% (Heaton et al., 1978). Im Gutachtenkontext wird bei Schädelhirntrauma-Patienten zum Teil ein inverser Zusammenhang zwischen dem Schweregrad und dem zeitlichen Verlauf der Beschwerden berichtet. Bei leichten Schädelhirntraumata sind häufiger negative Antwortverzerrungen zu erwarten als bei eindeutig objektivierbaren, ernsthaften zerebralen Störungen (Merten, 2009). In einer Untersuchung von Merten et al. (2006) mit neurologisch-psychiatrischen Gutachtenpatienten zeigten 51.1% der PTBS-Patienten, 45.6% der Patienten mit Gehirnprellung und 42.6% der Patienten mit Gehirnerschütterung Antwortverzerrungen. Die Prävalenzraten vorgetäuschter kognitiver Störungen bei Schmerzpatienten in Entschädigungsverhandlungen lagen nach Boone (2013) zwischen 17% (Gervais et al., 2004) und 42% (Richman et al., 2006), bei Begutachtungen in Folge toxischer Verletzungen zwischen 27.4% (van Hout, Schmand, Wekking & Deelman, 2006) und 40% (Greve et al., 2006). Mittenberg et al. (2002) fanden die höchsten Prävalenzen mit 41.2% bei Patienten mit Schädelhirntraumata, gefolgt von somatoformen Störungen (33.5%) und Fibromyalgie (38.6%) (Tabelle 4).

Tabelle 4

Adjustierte Prävalenzraten wahrscheinlicher Simulation / Aggravation einzelner Diagnosegruppen nach Mittenberg et al. (2002).

Diagnostische Kategorie	M % (95%-CI) adjustiert
Leichte Schädel-Hirn-Verletzung	41.2 (4.5)
Fibromyalgie / chronisches Fatiguesyndrom	38.6 (5.5)
Somatoforme Störungen inkl. Schmerzstörungen	33.5 (5.5)
Neurotoxische Störungen	29.5 (5.8)
Verletzungen durch Elektrizität	25.6 (5.5)
Depressive Störungen	16.1 (4.2)
Angststörungen	13.6 (4.3)
Dissoziative Störungen	10.5 (5.3)
Epilepsie	9.4 (3.4)
Mittel / -schwere Schädel-Hirn-Verletzung	8.8 (2.0)
Vaskuläre Demenz	2.4 (1.0)

Anmerkungen. Adjustierte Prävalenzraten mit korrigierten Varianzen verschiedener Bezugsquellen.

Die wichtigsten Untersuchungsdesigns der BV sind Analogstudien mit hoher interner und niedriger externer Validität, sowie *known-group Designs* mit hoher externer und niedriger interner Validität. Um die Häufigkeiten negativer Antwortverzerrungen für unterschiedliche Kontexte und Diagnosegruppen zu ermitteln, werden Prävalenzstudien durchgeführt. Die diagnostischen Schwierigkeiten der BV spiegeln sich auch in großen Spannweiten der Prävalenzraten wider. Mit steigender Anzahl an Prävalenzstudien kristallisieren sich allmählich vergleichbare Wertebereiche heraus und ergänzen die viel zitierte Expertenfrage von Mittenberg et al. (2002). Die höchsten Prävalenzraten von 30% bis 50% liegen im Gutachtenkontext, während im klinischen Kontext mit niedrigeren Prävalenzen von 5% bis 10% zu rechnen ist. Ferner werden bei Patienten mit leichten Schädelhirntraumata, neurotoxischen und somatoformen Störungen, einschließlich Schmerzstörungen, häufiger negative Antwortverzerrungen beobachtet als bei anderen Diagnosegruppen.

2.6 Erklärungsmodelle, kognitive Prozesse und neuronale Korrelate

Zum Verständnis der zugrunde liegenden Ursachen werden die Anreize simulativen Verhaltens sowie die Erklärungsansätze nach Rogers (1990a, 1990b) in diesem Abschnitt vertieft. Rogers entwickelte mit den pathogenetischen, kriminologischen und adaptiven

Modellen der Simulation drei Erklärungsansätze, die den deskriptiven Charakter der Klassifikationsmodelle ergänzen sollen. Um die Ätiologie simulativen Verhaltens zu verdeutlichen, werden auch die Analogien der kognitiven Prozesse des Lügens sowie die neuronalen Korrelate nicht-authentischer Verhaltensweisen dargestellt.

2.6.1 Erklärungsmodelle nach Rogers

Im pathogenetischen Modell wird eine Psychopathologie als Ursache der Simulation angenommen (Rogers, 1990a, 1990b). Das Vortäuschen von Symptomen wird als Kontrollversuch, im Sinne einer bewussten Reproduktion einer zugrunde liegenden unbewussten Psychopathologie, interpretiert. Allerdings können Simulation und Psychopathologie nicht automatisch gleichgesetzt werden, da nicht jeder Simulant zwangsläufig eine psychische Erkrankung hat. Obwohl das pathogenetische Modell keine ausreichende Erklärung darstellt (Rogers, 1990b), ist diese Ansicht in der Begutachtung kognitiver Beschwerden, auch unter Experten, heute noch verbreitet (Merten, 2014). Dies ist wahrscheinlich in den teilweise nur schwer nachvollziehbaren Verhaltensweisen zur Verdeutlichung vermeintlicher Rechtsansprüche begründet, die für einen Außenstehenden psychopathologisch erscheinen (Merten, 2014). Dennoch schließen sich Symptomsimulation und Psychopathologie nicht aus und es erscheint durchaus plausibel, dass Patienten bei dem Vortäuschen von Krankheiten von ihren Erfahrungen in psychiatrischen Kontexten Gebrauch machen (Rogers, 1990a).

Eine vorsätzlich antisoziale Motivation liegt dem kriminologischen oder puritanischen Modell zugrunde. Darin werden die simulativen Verhaltensweisen als Abweichung sozial akzeptierter Normen fokussiert und negativ bewertet. Simulanten werden als unbeeinträchtigte Personen betrachtet, die motiviert durch rechtliche oder ökonomische Anliegen das System hintergehen möchten (Merten, 2014; Rogers, 1990a, 1990b). Auch das DSM bezieht sich auf ein kriminologisches Erklärungsmodell und definiert die antisoziale Persönlichkeitsstörung als Diagnosekriterium einer Simulation. Auch die erhöhten Prävalenzraten für das Simulieren von Krankheitssymptomen im forensischen Kontext (22.8% laut Mittenberg et al. (2002)) untermauern einen Zusammenhang von Simulation und sozial devianter Verhaltensweisen.

Im adaptiven Modell wird Simulation als pragmatische Lösung im Umgang mit subjektiv ungünstigen Bedingungen gesehen. Das Modell basiert auf drei Annahmen: (1) Eine Person befindet sich unfreiwillig in einer Begutachtung oder Behandlung. (2) Die Person erwartet durch Selbstoffenbarung einen Verlust und durch Vortäuschen von Beschwerden einen Gewinn. (3) Sie betrachtet die Simulation von Krankheitssymptomen als effektivste Strategie für ihre Zielerreichung (Rogers, 1990a, 1990b; Rogers & Cavanaugh, 1983). Demzufolge ist

das adaptive Modell eine Erweiterung der Entscheidungstheorie. In einer risikoreichen Situation werden Entscheidungen entsprechend ihrer antizipierten Nützlichkeit und Wahrscheinlichkeit getroffen. Der Prozess wird als adaptiv bezeichnet, da durch Kosten-Nutzen-Abwägungen die effektivste Alternative zur Zielerreichung ausfindig gemacht wird (Rogers, 1990a). Empirisch gestützt wird das Modell durch den Zusammenhang zwischen den Beschwerdendarstellungen von Gefängnisinsassen und Veränderungen der situativen Bedingungen in einer Studie von Walters, White und Greene (1988). Die Häufigkeit von Aggravation bzw. Dissimulation variierte entsprechend der Vorteile (Einzelzelle) bzw. Nachteile (Strafmaß) einer psychischen Erkrankung. Vergleichbare Ergebnisse finden sich bei Audubon und Kirwin (1982) sowie Wasyliv, Grossman, Haywood und Cavanaugh (1988). Insgesamt erscheint das adaptive Modell aufgrund der plausiblen konzeptionellen Grundannahmen als Erklärungsmodell geeigneter als das pathogenetische und das kriminologische Modell, deren Dichotomie (krank / kriminell) es überwindet (Rogers, 1990a).

2.6.2 Anreize als Erklärung

Als Hauptanreize für Simulation gelten neben finanziellen Entschädigungen auch der Rückzug aus der Erwerbstätigkeit und die Vermeidung strafrechtlicher Konsequenzen (Nies & Sweet, 1994; Teichner & Wagner, 2004). Nies und Sweet (1994) sehen in Anlehnung an Miller und Cartlidge (1974) sowie Gilandas und Touyz (1982) einen Zusammenhang zwischen der Industrialisierung und der Zunahme an Arbeitsunfällen und dem Auftreten von Simulationen. Da Beeinträchtigungen durch Kompensationen beglichen werden, wird das Auftreten von Simulationen begünstigt. Binder und Rohling (1996) untersuchten in einer Metaanalyse mit 2353 Schädelhirntrauma-Patienten aus 18 Studien den Einfluss finanzieller Anreize auf die Symptomdarstellung in neuropsychologischen Testverfahren. Es zeigte sich ein inverser Zusammenhang zwischen monetären Zahlungen und dem Schweregrad der Beeinträchtigungen in neuropsychologischen Leistungsparametern. Patienten mit leichtem Schädelhirntrauma und Aussicht auf Kompensationszahlungen zeigten stärkere Leistungseinbußen als Patienten mit schwerem Schädelhirntrauma ohne Aussicht auf Kompensationszahlungen. Den Zusammenhang zwischen der Höhe monetärer Kompensationszahlungen und dem Ausmaß negativer Antwortverzerrungen untersuchten auch Bianchini et al. (2006) mit einer Stichprobe von 332 Schädelhirntrauma-Patienten. Neben einer Kontrollgruppe ohne Aussicht auf Kompensationszahlungen, konnte die Stichprobe in zwei weitere Gruppen entsprechend der Höhe der Entschädigungsansprüche (niedrig vs. hoch) unterteilt werden. Mit Zunahme der Kompensationssumme stieg auch die Anzahl negativer Antwortverzerrungen.

18% der Patienten mit niedrigen Entschädigungsansprüchen und 27% der Patienten mit hohen Entschädigungsansprüchen erfüllten die Slick-Kriterien einer wahrscheinlichen SND.

2.6.3 Kognitive und neuronale Prozesse des Lügens

Um die Ätiologie der Simulation kognitiver Defizite zu erklären, stellte Seron (2014) in seiner Überblicksarbeit den Zusammenhang zwischen Lügen und Simulation dar und begründete diesen in der Intentionalität beider Prozesse. Der Autor definierte Lügen nach Mahon (2008) als die Kommunikation einer bewusst falschen Aussage mit der Absicht, dass eine andere Person diese glaubt und für wahr hält. Lügen werden als Teil unseres Alltags betrachtet. DePaulo, Kashy, Kirkendol, Wyer und Epstein (1996) fanden beispielsweise heraus, dass gesunde Personen durchschnittlich ein bis zwei Mal am Tag lügen. Mit vier Jahren beginnt der Mensch erste Unwahrheiten zu erzählen (Stern, 1965). Nach der *Theory of mind* stellen die Fähigkeit zur Perspektivenübernahme, Antizipation der Reaktion des Interaktionspartners sowie Differenzierung zwischen subjektiver Überzeugung und Realität die wesentlichen Voraussetzungen des Lügens dar (Jager, 2012). Lügen und analog das Vortäuschen von Symptomen kann somit als kognitiv anspruchsvoller Prozess angesehen werden. Neben der Antizipation, was der andere denkt und hören möchte, erfordert es auch ein hohes Maß an Emotionsregulation, um keine Nervosität oder Unsicherheit zu zeigen. Insofern bezeichnet Jager (2012) eine „Lüge als Leistung“ (Jager, 2012; S. 282., Z. 19). Zudem werden neben der Entscheidung zu Lügen auch Inhibitionsfunktionen zum Unterdrücken der Wahrheit, planerisches Denken sowie kognitive Flexibilität bei der Umsetzung der Lüge benötigt. Somit beansprucht Lügen insbesondere die exekutiven Funktionen (Seron, 2014). Darüber hinaus werden mnestischen Funktionen benötigt, da man sich zur Aufrechterhaltung der Lüge stets bewusst an deren Details erinnern muss. Die spezifischen Verhaltensweisen, die mit den kognitiven Prozessen des Lügens einhergehen, sind mit der Bearbeitung komplexer, anspruchsvoller Aufgaben vergleichbar. Es wird mehr Zeit beansprucht eine Lüge zu generieren, als wahrheitsgemäß zu antworten (Seron, 2014). Dies führt zu verzögerten Reaktionszeiten und erhöhten Fehlerraten in kognitiven Leistungstests (Seymour & Kerlin, 2008; Sheridan & Flowers, 2010). Darüber hinaus werden unzutreffende Angaben subjektiv anstrengender erlebt als wahrheitsgemäße Aussagen (Schagen, Schmand, Sterke & Lindeboom, 1997).

Zur Differenzierung der zugrunde liegenden kognitiven Funktionen des Lügens analysierte Seron (2014) die bisherige Forschungsliteratur. Er verdeutlichte, dass bereits durch das Lügen selbst kognitive Ressourcen beansprucht werden, die dann nicht mehr für die

Aufgabenbearbeitung zur Verfügung stehen. Auf neuronaler Ebene wird angenommen, dass beim Lügen ein ausgedehntes Netzwerk aktiviert wird, ähnlich wie bei Anforderungen an diverse Exekutivfunktionen. Laut Seron (2014) zeigten sich vor allem Aktivierungen im präfrontalen Cortex (einschließlich des orbito-frontalen / ventrolateralen präfrontalen Cortex, des dorsolateralen präfrontalen Cortex), anterioren Cingulum, Thalamus, Temporal- und Parietallappen, Caudatum und in der Insula.

2.6.4 Neuronale Korrelate in bildgebenden Verfahren

Die Verwendung bildgebender Verfahren (*Neuroimaging*) in der BV-Forschung soll dazu beitragen, die zerebralen Aktivitäten, die mit einer invaliden Symptomdarstellung einhergehen, zu identifizieren und neuronale Korrelate abzuleiten (Bigler, 2015). Durch die Visualisierung der involvierten Hirnstrukturen während der Bearbeitung von PVT sollen bisherige Annahmen überprüft und ergänzt werden. Dabei wird die Anstrengungsbereitschaft, analog zu Gedächtnis oder Aufmerksamkeit, als kognitive Funktion betrachtet, die sensorisch-perzeptuelle und attentionale Prozesse aktiviert und die entsprechenden neuronalen Systeme beansprucht (Larsen, Allen, Bigler, Goodrich-Hunsaker & Hopkins, 2010). Bigler (2015) verdeutlichte diese Annahme mit Hilfe der Verarbeitungsmodi von Braver et al. (2014) und Pessoa und Engelmann (2010), die das Zusammenspiel von Anstrengungsbereitschaft (Motivation) und Aufmerksamkeit bei der Verhaltenssteuerung (Aufgabenbearbeitung) beschreiben. Die Verarbeitung zahlreicher motivationaler und attentionaler Prozesse kann parallel verlaufen oder sich wechselseitig beeinflussen (Integration). Zudem können die attentionalen Prozesse einer Aufgabenbearbeitung durch motivationale Systeme beeinflusst werden (Mediation). Somit ist von einer komplexen neuronalen Aktivierung auszugehen.

Während in der BV-Literatur bislang die Annahme vertreten wurde, dass bei der authentischen Bearbeitung von PVT nur wenig kognitive Ressourcen beansprucht werden, konnte dies durch Untersuchungen mit bildgebenden Verfahren nicht bestätigt werden. Vielmehr zeigten sich vergleichbare Aktivitätsmuster wie bei kognitiv anspruchsvollen Aufgaben. Beispielsweise verdeutlichte die funktionelle Magnetresonanztomographie (fMRT)-Studie von Allen, Bigler, Larsen, Goodrich-Hunsaker und Hopkins (2007), dass der authentischen Bearbeitung des *Word Memory Tests* (*WMT*: Green, 2003, Revised 2005) ein komplexes neuronales Netzwerk zu Grunde liegt. Neben dem dorsolateralen präfrontalen Kortex wurden auch der superiore Parietallappen, das anteriore Cingulum, die anteriore Insula und das frontale Operculum aktiviert. Die Ergebnisse verdeutlichen, dass der Bearbeitung des *WMT* komplexe neuronale Netzwerke zugrunde liegen und ein beträchtliches Maß an Aufmerksamkeit, Konzentration

und Arbeitsgedächtnis beanspruchen. Ähnliche fMRT-Ergebnisse fanden Larsen et al. (2010) bei der Bearbeitung des *WMT* von zehn klinisch unauffälligen Probanden mit verminderter als auch uneingeschränkter Anstrengungsbereitschaft. Während des verzögerten Abrufs des *WMT* zeigten sich die ausgeprägtesten neuronalen Aktivitäten im dorsolateralen präfrontalen Kortex, superioren Parietallappen, anterioren Cingulum, frontalen Operculum und in den bilateralen lingualen Kortizes. Die Aufgabenbearbeitung mit verminderter Anstrengungsbereitschaft beanspruchte zudem weitere kortikale Areale, wie die rechts anteriore Insula und die bilateralen lingualen Kortizes.

In einer weiteren Bildgebungsstudie von Kosheleva, Spadoni, Strigo, Buchsbaum und Simmons (2016) mit klinisch unauffälligen Probanden ($N = 22$) zeigten sich stärkere zerebrale Aktivitäten bei den Probanden, die einen PVT mit verminderter Anstrengungsbereitschaft bearbeiteten als bei den Probanden mit uneingeschränkter Anstrengungsbereitschaft. Interessanterweise zeigten die Probanden mit verminderter Anstrengungsbereitschaft eine stärkere zerebrale Aktivierung bei korrekter als bei falscher Bearbeitung des *Test of Memory Malingering (TOMM: Tombaugh, 1996)*. Im fMRT wurde eine stärkere Aktivität im superioren und medialen präfrontalen Kortex der Probanden mit verminderter Anstrengungsbereitschaft gemessen. Diese Ergebnisse verdeutlichen, dass die Präsentation vorgetäuschter Gedächtnisdefizite und vor allen die korrekten Antworten bei nicht-authentischer Aufgabenbearbeitung mit einer stärkeren zerebralen Aktivierung, im Sinne der kognitiven Kontrolle, einhergeht. Zudem wird die substanzielle Ähnlichkeit zwischen den neuronalen Korrelaten bei der anspruchsvollen Fehlerpräsentation (*effortful failure*) und der anspruchsvollen Leistungspräsentation (*effortful success*) in PVT verdeutlicht (Spadoni, Kosheleva, Buchsbaum & Simmons, 2015).

Insgesamt ermöglichen bildgebende Verfahren einen weiteren Zugang zur Differenzierung zwischen tatsächlichen kognitiven Beeinträchtigungen und verminderter Anstrengungsbereitschaft. Neben verminderter Anstrengungsbereitschaft können auffällige Ergebnisse in PVT jedoch auch auf neurobiologische Dysfunktionen durch Hirnschädigungen zurückzuführen sein (Larsen et al., 2010). Beispielsweise zeigten in einer Studie von Spadoni et al. (2015) 10 von 45 Kriegsveteranen mit PTBS oder leichtem Schädelhirntrauma und ohne objektivierbaren Anreiz zur Symptomverzerrung eine auffällige Testperformanz im *Test of Memory Malingering (TOMM: Tombaugh, 1996)*. Im PET wurde bei den Patienten mit auffälliger Testperformanz ein geringerer Metabolismus im ventromedialen präfrontalen Kortex als bei Patienten mit unauffälliger Testperformanz deutlich. Anhand des Metabolismus

im PET konnten die auffälligen PVT-Ergebnisse auf die kortikalen Dysfunktionen der PTBS oder des Schädelhirntraumas zurückgeführt und Falsch-Positiv-Klassifikationen auf Grund auffälliger Testwerte vermieden werden. Auch Bigler (2015) betont den diagnostischen Zugewinn durch bildgebende Verfahren im Rahmen der BV.

Rogers (1990a, 1990b) entwickelte drei Erklärungsmodelle für die Ursachen simulativer Verhaltensweisen. Auf Grund der Berücksichtigung situativer Faktoren erscheint das adaptive Modell plausibler als das pathologische und kriminalistische Modell. Darüber hinaus stellen finanzielle Anreize ein bedeutsames Motiv für negative Antwortverzerrungen dar. Wie beim Lügen werden bei negativen Antwortverzerrungen in neuropsychologischen Untersuchungen sowohl attentionale, exekutive und mnestiche Funktionen als auch emotionale Regulationsmechanismen beansprucht. Die Visualisierung dieser komplexen neuronalen Netzwerke anhand bildgebender Verfahren zeigt eine hohe Aktivität im präfrontalen Kortex bei der Bearbeitung von PVT, sowohl bei verminderter als auch unbeeinträchtigter Anstrengungsbereitschaft. Zudem werden Areale im Parietallappen, Cingulum und Operculum aktiviert.

2.7 Negative Antwortverzerrungen in neuropsychologischen Untersuchungen

In diesem Kapitel werden die unterschiedlichen diagnostischen Vorgehensweisen zur Erfassung negativer Antwortverzerrungen in neuropsychologischen Untersuchungen mit Schwerpunkt der diagnostischen Methoden dieser Arbeit dargestellt.

2.7.1 Beeinflussbarkeit negativer Antwortverzerrungen

Probanden erzielen negative Antwortverzerrungen in neuropsychologischen Untersuchungen über Formen der Leistungsverweigerung. Dazu gehören neben einer erhöhten Fehleranzahl und verlangsamten Reaktionszeiten auch das Zurückhalten korrekter Antworten, wobei die Darstellung der Leistungsdefizite durch die Schwierigkeit der Testaufgaben beeinflusst werden kann (Steck, Reuter, Meir-Korrell & Schönle, 2000). Auch Rees, Tombaugh, Gansler und Moczynski (1998) gehen davon aus, dass das Vortäuschen kognitiver Beeinträchtigungen mehr als die bloße Absicht dazu erfordert. Vielmehr hängt die Symptompräsentation vom subjektiven Krankheitsverständnis ab und inwieweit dieses in ein glaubwürdiges neuropsychologisches Leistungsprofil übersetzt werden kann. Vielen Simulanten fehlt jedoch Fachwissen über kognitive Defizite und deren neuropsychologische Korrelate. Darüber hinaus beeinflusst nach Bolan, Foster, Schmand und Bolan (2002) auch das kognitive

Ausgangsniveau die Simulationsleistung und die Autoren verweisen auf den positiven Zusammenhang zwischen Ausbildungsstatus und instruierter Simulationsleistung, den Schwartz, Gramling, Kerr und Morin (1998) ermittelten. Demnach gehen mit höheren Bildungsabschlüssen auch glaubhaftere Simulationsleistungen einher.

Neben der Intention zur Symptompräsentation, dem kognitiven Ausgangsniveau und dem störungsspezifischen Vorwissen können auch Informationen, die aus der Untersuchung selbst resultieren, die präsentierte Testperformanz beeinflussen. Reicker (2008) geht davon aus, dass Personen mit Simulationsintention besonders auf Informationen achten, die sie aus der neuropsychologischen Untersuchungssituation gewinnen können. In diesem Zusammenhang betont Bernard (1990) den Einfluss der Reihenfolge der Testverfahren in einer neuropsychologischen Untersuchung. Die Testadministration schwieriger Verfahren zu Beginn der Untersuchung kann eine Sensibilisierung gegenüber leichteren darauffolgenden PVT begünstigen. In einer Analogstudie variierte Reicker (2008) die Reihenfolge der Aufgaben der *Computerized Tests of Information Processing (CTIP: Tombaugh & Rees, 2008)*, in dem die Verfahren in aufsteigender bzw. absteigender Komplexität angeordnet wurden. Allerdings zeigte sich kein bedeutsamer Einfluss der Reihenfolge sowie der Anzahl der Testverfahren. Laut einer Umfrage von Slick et al. (2004) werden in der klinischen Praxis PVT etwa in der Hälfte aller Fälle immer zu Beginn der Untersuchung durchgeführt, während 45.8% der Experten einen PVT erst im Verlauf der Untersuchung bei entsprechendem Verdacht anwenden.

Eine weitere Einflussgröße stellt nach Bolan et al. (2002) die Rückmeldung der Testleistungen dar. Das Feedback kann als Hilfestellung zur Einschätzung der eigenen Testperformanz fungieren und dadurch das weitere Simulationsverhalten nach den Vorstellungen über das Ausmaß der gewünschten Beeinträchtigungen beeinflussen (Tombaugh, 1997). Dementsprechend befürworten Prigatano, Smason, Lamb und Bortz (1997) das Zurückhalten von Rückmeldungen, um Probanden den Überblick über ihr Leistungsverhalten nicht zu erleichtern. Andererseits können Rückmeldungen über die Richtigkeit der dichotomen Antworten in PVT bei Personen mit Simulationsintention auch zur Einschätzung ihrer Testleistung als zu gut führen und zur absichtlichen fehlerhaften Testperformanz ermutigen. In der Analogstudie von Bolan et al. (2002) hatte die Rückmeldung über die Testleistung jedoch keinen inferenzstatistisch bedeutsamen Einfluss auf die Testperformanz. Dennoch betonen die Autoren den bedachten Einsatz von Rückmeldungen im Einzelfall in der klinischen Praxis.

2.7.2 Erfassung negativer Antwortverzerrungen

Als Grundlage der BV gelten das Überprüfen der präsentierten Symptome hinsichtlich ihrer Plausibilität und der geschilderten Beschwerden hinsichtlich ihrer Konsistenz (Merten, 2014; Merten, Stevens, et al., 2007). Die Analyse der Diskrepanzen zwischen beobachtbaren und prototypischen Symptomausprägungen der entsprechenden Erkrankung wird von Larrabee et al. (2007) als *Diskrepanzmethode (discrepancy method)* bezeichnet. Dabei wächst die Überzeugung des Klinikers, es mit einer nicht-authentischen Diagnose zu tun zu haben, je ausgeprägter diese subjektiv bewertete Diskrepanz ist. Jedoch warnt Miller (2001) davor, seltene und ungewöhnliche Symptome aus Unwissenheit als psychogen oder vorgetäuscht zu klassifizieren. Auf Grund der potentiell beträchtlichen Auswirkungen einer Fehlklassifikation für den Patienten muss in der klinischen Praxis das Risiko von Falsch-Positiv-Entscheidungen minimiert werden (Larrabee, 2014). Deswegen wird beim Einsatz psychometrischer Methoden zur BV eine niedrigere Sensitivität zugunsten einer höheren Spezifität forciert. Während die Anzahl an Falsch-Positiv-Klassifikationen so gering wie möglich gehalten wird, bleiben dadurch Personen mit negativen Antwortverzerrungen möglicherweise unentdeckt. Laut einer Metaanalyse von Vickery, Berry, Inman, Harris und Orey (2001) liegt der Anteil an Falsch-Negativ-Klassifikationen in Stichproben mit Simulationspatienten im Durchschnitt bei 44%. Demzufolge empfehlen Bender und Rogers (2004) sowie Bush et al. (2005) nicht nur die Anwendung mehrerer PVT (*multi-test*), sondern auch verschiedener Methoden (*multi-method*).

2.7.3 Diagnostische Ansätze

Rogers, Harrell und Liff (1993) gliedern die Methoden der BV in sechs diagnostische Ansätze, auf die häufig Bezug genommen wird (Bender & Rogers, 2004; Merten, 2002, 2009, 2014; Merten, Stevens, et al., 2007):

1. Testdeckeneffekt und Testbodeneffekt
2. Leistungskurve
3. Fehlergröße
4. Alternativwahlverfahren
5. Inkonsistente oder untypische Leistungsprofile
6. Präsentation psychischer Folgen

1. Testdecken- und Testbodeneffekt

Es ist anzunehmen, dass Aufgaben mit objektiv geringem Schwierigkeitsgrad eine Einladung zur fehlerhaften Bearbeitung für Personen mit Simulationsintention darstellen. Somit werden Aufgaben falsch beantwortet, die selbst von Personen mit ausgeprägten Pathologien korrekt beantwortet werden können, wie z.B. „Was ist größer: ein Hund oder ein Pferd?“ (Rogers et al., 1993). Merten (2014) bezeichnet dieses methodische Vorgehen als Testdeckeneffekt. Dieses Prinzip findet häufig Verwendung in PVT (Merten, 2002). Durch Instruktionshinweise einer vermeintlichen Schwierigkeit der Aufgaben wird oftmals versucht, über deren Einfachheit hinwegzutäuschen. Darin sehen Rogers et al. (1993) ein ethisches Problem, welches bisher nicht gelöst worden ist.

2. Leistungskurve

Die kognitive Leistungsfähigkeit einer Person wird anhand einer neuropsychologischen Leistungskurve der zugrunde liegenden Funktionsbereiche abgebildet. Die Interpretation von Leistungskurven in der Performanzvalidierung ist auf Goldstein (1945) zurückzuführen. Er deutete als erster die Diskrepanz zwischen Fehlern bei einfachen Aufgaben und gleichzeitig korrekt bearbeiteten weit anspruchsvolleren Aufgaben als Hinweis für negative Antwortverzerrungen (Rogers et al., 1993). Es wird angenommen, dass sich Simulanten bei der Aufgabenbearbeitung auf die Darstellung kognitiver Beeinträchtigungen fokussieren und den Schwierigkeitsgrad der Aufgaben dabei nicht berücksichtigen. Darüber hinaus können störungsspezifische Leistungskurven von bona-fide Patienten oder prototypische Verläufe in bestimmten Funktionsbereichen wie der Gedächtnis- und Aufmerksamkeitsfunktionen herangezogen und als Referenzwerte betrachtet werden (Rogers et al., 1993). Die Leistungskurven im Funktionsbereich Gedächtnis weisen in der Regel bessere Leistungen im Wiedererkennen als im Abruf auf. Hinsichtlich der Abrufleistungen gelingt der gestützte Abruf mit Abrufhilfe gewöhnlich besser als der freie Abruf (Merten, 2002, 2014). Grundsätzlich ist mit Zunahme der Gedächtnisstützen eine Verbesserung der Gedächtnisleistung zu erwarten. Im Funktionsbereich der Aufmerksamkeit zeigt sich in Reaktionszeitanalysen typischerweise eine Verlangsamung der Reaktionszeiten bei steigender Aufgabenkomplexität (Merten, 2002). Hinsichtlich basaler Aufmerksamkeitsfunktionen sind schnellere Reaktionszeiten als bei komplexeren Aufmerksamkeitsfunktionen zu erwarten (Merten, 2014).

Die Analyse von Leistungskurven gilt als plausible Methode zur Erfassung negativer Antwortverzerrungen. Allerdings können Schwankungen in den Leistungskurven durch

zahlreiche weitere Ursachen (Ermüdung, Veränderungen im psychischen Status, hirnorganische Schädigungen) hervorgerufen werden (Merten, 2002). Merten (2014) empfiehlt eine gewisse Fehlervarianz zu berücksichtigen, um Falsch-Positiv-Entscheidungen zu vermeiden.

3. Fehlergröße

Der Ansatz der *Fehlergröße* (*Magnitude of Error*) bezieht sich auf die quantitative und qualitative Analyse falscher Antworten während der Bearbeitung neuropsychologischer Standardverfahren. Als typisch für intendierte Antwortverzerrungen gelten *Knapp-daneben-Antworten* (Rogers et al., 1993), wie beispielsweise trotz Instruktionsverständnis bei Reaktionszeitaufgaben stets auf den falschen Stimulus zu reagieren. Somit sind Knapp-daneben-Antworten durch ein regelhaftes, sehr geringfügiges Abweichen von der korrekten Lösung gekennzeichnet. Auf Grund minimaler, jedoch zielgerechter Abweichungen von der korrekten Lösung sprechen diese Antwortmuster für deren Kenntnis und eine systematische Antwortmanipulation (Merten, 2002, 2014). Differentialdiagnostisch ist das Ganser-Syndrom von Knapp-daneben-Antworten einer Simulation abzugrenzen. Das Ganser-Syndrom ist durch das Vorbeiantworten simpler Fragen (bspw. „2 plus 2 ergibt 5“) charakterisiert (Merten, 2014; Rogers et al., 1993) und wird häufig von dissoziativen Phänomenen begleitet (Ebner & Kopp, 2014). Als Voraussetzung für eine standardisierte Erfassung der Fehlergröße werden Testverfahren benötigt, die über diverse Antwortwahlmöglichkeiten mit einer Wahrscheinlichkeitsannahme für jede Antwortselektion verfügen (Martin, Franzen & Orey, 1998). In einer Studie von Martin et al. (1998) wählten sowohl experimentell instruierte Simulanten als auch Patienten mit Simulationsverdacht signifikant häufiger unwahrscheinliche Items aus und erzielten eine höhere Fehlergröße. Insgesamt liegen erst wenige empirische Untersuchungen vor, so dass keine ausreichenden Kriterien und Einschätzungen über die Häufigkeit von Knapp-daneben-Antworten bestehen (Merten, 2002). Dennoch profitiert die einzelfallbezogene, qualitative Fehleranalyse laut Merten (2014) von dieser methodischen Vorgehensweise.

4. Alternativwahlverfahren

Alternativwahlverfahren sind speziell entwickelte Testverfahren zur Beurteilung der Validität von Testprofilen und ermöglichen Rückschlüsse auf die Beschwerdvalidität geschilderter kognitiver Beeinträchtigungen (Blaskewitz & Merten, 2006). Für Alternativwahlverfahren wurde in der Forschungsliteratur zunächst die Bezeichnung Beschwerdvalidierungstests

verwendet. Seit der Differenzierung von Larrabee (2012) werden sie meist als Performanzvaliditätstests (PVT) verwendet. Sie gelten als der bedeutsamste diagnostische Ansatz in der BV (Merten, 2005, 2014) und werden deshalb ausführlich in Kapitel 2.8 beschrieben.

5. Inkonsistente oder untypische Leistungsprofile

Schwankungen der Testperformanz innerhalb eines Funktionsbereichs, die mittels mehrerer Verfahren oder wiederholt erfasst werden, deuten durch inkonsistente oder untypische Leistungsprofile auf negative Antwortverzerrungen (Merten, 2014; Rogers et al., 1993). Diskrepanzen zwischen Ergebnissen, die nicht stimmig sind oder den typischen kognitiven Leistungsprofilen der diagnostizierten Erkrankungen widersprechen, ergeben „neuropsychologisch keinen Sinn“ (Merten & Dohrenbusch, 2010, S. 398) und deuten auf negative Antwortverzerrungen hin. Laut Cullum, Heaton und Grant (1991) ist die Voraussetzung einer konsistenten Antwortmanipulation die Produktion äquivalenter Beeinträchtigungen über mehrere Testverfahren und Zeitpunkte. Dies erfordert ein hohes Maß an Selbstüberwachung (*self-monitoring*) und ein elaboriertes Wissen über menschliche Verhaltensweisen, worüber die meisten Simulanten laut Strauss et al. (2002) nicht verfügen. Folglich eignet sich die serielle Verwendung längerer standardisierter Untersuchungsbatterien, wie beispielsweise die *Wechsler Adult Intelligence Scale-IV* (*WAIS-IV*: Wechsler, 2008) oder die *Wechsler Memory Scale-IV* (*WMS-IV*: Wechsler, 2009) zur Analyse inkonsistenter Testprofile als aussagekräftige Methode. Strauss et al. (2002) untersuchten die intraindividuelle Variabilität der Testperformanz ihrer Probanden an drei testpsychologischen Messzeitpunkten. Im Vergleich zur Kontrollgruppe präsentierten die experimentellen Simulanten eine schlechtere Testperformanz und eine größere intraindividuelle Variabilität über die Messzeitpunkte hinweg. Die Simulanten zeigten zudem stärkere Fluktuationen der Bearbeitungsgenauigkeit bei schwierigen als bei einfachen Aufgaben.

Insgesamt bedarf dieser Ansatz einer sehr umsichtigen Interpretation, da Inkonsistenzen und untypische Testprofile häufig bei Hirnverletzungen und psychischen Erkrankungen zu beobachten sind. Zudem ist er aufwendig, zeitintensiv und eignet sich dadurch nur bedingt für einen routinemäßigen Einsatz in der klinischen Praxis (Pankratz, 1988).

6. Präsentation psychischer Folgen

Im Rahmen der Symptomvalidierung werden psychische Symptome hinsichtlich ihrer Plausibilität des entsprechenden Störungsbildes überprüft. Psychische Symptome können

jedoch nur schwer objektiviert werden, da sie selten durch direkt messbare Parameter abgebildet werden können. Zudem kommen Symptome wie beispielsweise Kopfschmerzen in der Allgemeinbevölkerung häufig vor und haben an sich eine hohe Prävalenz (Merten, 2002). Des Weiteren können diese Beschwerden tatsächlich als Begleiterscheinungen von Hirnschädigungen oder deren Folgen auftreten. Zudem können sie bereits vor der Erkrankung bestanden haben und dadurch eine eindeutige Zuordnung erschweren (Merten, 2014; Rogers et al., 1993). Zur Symptomvalidierung psychischer Beschwerden werden Validitäts- und Kontrollskalen in Fragebögen verwendet, die als *Selbstauskunfts-Validitätstests (SRVT)* bezeichnet werden (Larrabee, 2012; Merten, Friedel, Mehren & Stevens, 2007; Merten et al., 2016). Durch die standardisierten Selbstauskünfte kann das diagnostische Vorgehen ergänzt und validiert werden (Merten, 2010). Im englischsprachigen Raum kommen auch Fremdbeurteilungsverfahren wie beispielsweise das *Miller-Forensic Assessment of Symptoms Test (M-FAST)*: Miller, 2001) zum Einsatz, wofür aktuell keine entsprechenden deutschsprachigen Verfahren vorliegen.

Im deutschen Sprachraum wird laut der Expertenbefragung von Dandachi-FitzGerald et al. (2013) mit 24% am häufigsten die deutsche Adaptation des *Minnesota Multiphasic Personality Inventory (MMPI-2)*: Engel, 2000) verwendet, gefolgt von dem *Strukturierten Fragebogen Simulierter Symptome (SFSS)*: Cima, Merckelbach, Hollnack & Knauer, 2003) mit 18% (Dandachi-FitzGerald et al., 2015). Das *Minnesota Multiphasic Personality Inventory* dient mit 567 Items der Erfassung von Persönlichkeitsstrukturen im klinischen Kontext. Anhand von Kontrollskalen (L Lügenskala; F Seltenheitsskala, Simulation; K Dissimulation) sowie von Indexwerten (Index F-K) und einer *Symptom Validity Scale*, die eine Weiterentwicklung der *Fake Bad Scale* von Lees-Haley, English und Glenn (1991) ist, kann die Validität des Persönlichkeitsprofils überprüft werden. Der *Strukturierte Fragebogen Simulierter Symptome (SFSS)* stellt die deutsche Version des *Structured Inventory of Malingered Symptomatology (SIMS)*: Widows & Smith, 2005) dar. Es werden atypische und seltsame Symptome für die fünf Krankheitsbilder Psychose, affektive Störung, neurologische Beeinträchtigungen, amnestische Störung und niedrige Intelligenz anhand von 75 Items erfragt. Die Glaubhaftigkeit der Aufgabenbearbeitung wird dabei mit Hilfe von Schwellenwertvergleichen überprüft. Auch der störungsspezifische *Morel Emotional Numbing Test* (Morel, 2012) für posttraumatische Belastungsstörungen und emotionale Taubheit findet laut Merten (2014) zunehmend Verwendung im deutschen Sprachraum. Ein weiterer Selbstauskunfts-Validitätstest ist das *Verhaltens- und Erlebensinventar (VEI)*: Engel

& Groves, 2012). Auch die deutschsprachige Adaptation des *Personality Assessment Inventory (PAI)* von Morey (1991) beinhaltet 4 Validitätsskalen (Inkonsistenz, Seltenheit, ungünstiger Eindruck, günstiger Eindruck) und zwei zusätzliche Validitätsindikatoren (Simulationsindex, Rogers Diskriminanzfunktion).

Ein neueres Verfahren stellt das *Self-Report Symptom Inventory (SRSI)* dar, das 2016 von Merten, Merckelbach, Giger und Stevens entwickelt wurde. Dieser Fragebogen beinhaltet neben fünf Validitätsskalen zu unterschiedlichen Pseudosymptomen (Kognition/ Gedächtnis, neurologische Motorik, neurologische Sensorik, Schmerz, Angst/ Depression mit PTBS) fünf weitere Skalen zur Erfassung tatsächlicher klinischer Symptome (Kognition, Depression, Schmerz, unspezifisch somatisch, PTBS/ Angst). Die Validierungsstudie mit einer forensischen Stichprobe ($n = 387$) zeigte signifikante Interkorrelationen zwischen dem *Self-Report Symptom Inventory* und dem *Strukturierten Fragebogen Simulierter Symptome* ($r \geq .80$). Zudem wurde ein inverser Zusammenhang zwischen verminderter Anstrengungsbereitschaft im *Word Memory Test (WMT)* und verstärkter Darstellung von Pseudosymptomen ermittelt. Die interne Konsistenz des Verfahrens lag bei Cronbach's $\alpha > .90$ und die Retest-Reliabilität bei $r_{tt} > .85$. Die Klassifikationsgüte wurde als hervorragend bewertet ($AUC = .93$), so dass dieser Selbstauskunfts-Validitätstest ein vielversprechendes weiteres Verfahren zur Symptomvalidierung darstellt.

Negative Antwortverzerrungen können in neuropsychologischen Untersuchungen durch zahlreiche Faktoren wie das kognitive Ausgangsniveau, Vorwissen, Rückmeldungen über die Testperformanz oder die Aufgabenschwierigkeit beeinflusst werden. Von den sechs diagnostischen Vorgehensweisen zur Überprüfung der Konsistenz und Plausibilität präsentierter Symptome hat sich empirisch und klinisch insbesondere die Verwendung von Performanzvalidierungstests (PVT) und Selbstauskunfts-Validitätstests (SRVT) bestätigt. Darüber hinaus werden die Analysen von Testdeckeneffekten und Testbodeneffekten sowie Leistungskurven in die Beurteilung der Performanzvalidität einbezogen. Die Analyse der Fehlergröße und inkonsistenter Leistungsprofile gilt als aufwendig und empirisch wenig untersucht, kann allerdings zur qualitativen Einzelfallanalyse herangezogen werden.

2.8 Performanzvaliditätstests (PVT)

In diesem Abschnitt werden die Konstruktionsprinzipien, Erscheinungsformen, Gütekriterien und Interpretation der PVT erläutert. Rey entwickelte 1958 mit dem *15-Item-Test* den ersten,

standardisierten PVT, der bis heute Verwendung findet (Merten, 2014). Ende der 1980er Jahre begann im angloamerikanischen Raum mit dem Einsatz der PVT die Entwicklung der modernen BV (Merten et al., 2013). Dies zeigte sich auch in der zunehmenden Verwendung von PVT in der klinischen Praxis. Während 2007 63.6% der befragten nordamerikanischen Neuropsychologen PVT verwendeten, waren es 2015 bereits 91.8% (Martin et al., 2015). Eine Umfrage deutschsprachiger neuropsychologischer Sachverständiger im Jahr 2012 ergab, dass 61.3% im gutachterlichen und 13.2% im klinischen Kontext PVT immer oder in der Mehrzahl der Fälle anwendeten (Dandachi-FitzGerald et al., 2015).

2.8.1 Konstruktionsprinzipien

PVT basieren auf der Annahme, dass ihre Bearbeitung weitgehend unabhängig von der tatsächlichen kognitiven Leistungsfähigkeit erfolgt (Willis, Farrer & Bigler, 2011). Sie sind nach dem *Prinzip der verdeckten Leichtigkeit* konstruiert. Die Bearbeitung eines PVT erscheint demnach kognitiv anspruchsvoller, als es dem tatsächlichen Schwierigkeitsgrad der Aufgaben entspricht (Bolan et al., 2002; Rüsseler, Brett, Klaue, Sailer & Münte, 2008). Methodisch gesehen provozieren PVT bei unverminderter Anstrengungsbereitschaft Deckeneffekte in den Testergebnissen (Eberl & Wilhelm, 2007). Auf Grund ihrer geringen Schwierigkeit können PVT in der Regel von Patienten mit psychiatrischen, neurologischen oder leichten kognitiven Störungen korrekt gelöst werden (Merten & Dohrenbusch, 2010). Sobald ein PVT jedoch als zu einfach für eine glaubhafte Darstellung kognitiver Beeinträchtigungen erscheint, verliert das Verfahren an Sensitivität (Guilmette, Hart & Giuliano, 1993). Ein PVT muss demzufolge als anspruchsvolles Verfahren wahrgenommen werden, das für Patienten mit kognitiven Beeinträchtigungen entwickelt wurde (Reicker, 2008).

Ein weiteres Konstruktionsprinzip stellt das bereits erwähnte *Alternativwahlverfahren* dar. Die Verwendung von Alternativwahlverfahren ist Loren Pankratz (1979) zuzuschreiben, der die Methode zunächst zur Erfassung vorgetäuschter sensorischer Beeinträchtigungen (Taubheit) anwendete. Dabei sollten Patienten in einer Aufgabe mit zwei Antwortalternativen (ja oder nein) angeben, ob sie einen Ton gehört hatten oder nicht. Durch die dichotome Aufgabenstruktur (*forced-choice*) muss sich der Patient für eine der beiden vorgegeben Antworten entscheiden (Eberl & Wilhelm, 2007). In der Literatur findet man deswegen auch die Bezeichnung *Zwangswahlverfahren* (Blaskewitz & Merten, 2007). Bei jedem dichotomen Testitem besteht eine Zufallswahrscheinlichkeit von 50% für die korrekte Itemauswahl. Alternativwahlverfahren basieren somit auf stochastischen Wahrscheinlichkeitsrechnungen

und das tatsächliche Antwortverhalten eines Probanden wird mit der Ratewahrscheinlichkeit in Relation gesetzt (Boone, 2013). Ein Testergebnis unterhalb der Ratewahrscheinlichkeit ist äußerst unwahrscheinlich, so dass von einer bewussten Antwortmanipulation ausgegangen werden kann (Eberl & Wilhelm, 2007). In der klinischen Praxis erzielen weniger als 15% der Patienten mit nicht-authentischer Symptompräsentation ein Ergebnis signifikant unterhalb der Ratewahrscheinlichkeit (Boone, 2013). Demzufolge kommt der Ermittlung verfahrensspezifischer Schwellenwerte besondere Bedeutung zu. Diese empirischen Grenzwerte definieren, ab welchem Wert die Validität der präsentierten Testleistung für den jeweiligen PVT als eingeschränkt zu betrachten ist. Die Grenzwerte werden aus Sensitivitäts- und Spezifitätsanalysen abgeleitet und haben mindestens eine Spezifität von 90%, um die Rate an Falsch-Positiv-Fehlklassifikationen zu minimieren (Boone, 2013).

Das wachsende Forschungsinteresse geht mit einer steigenden Anzahl an PVT einher, die in der Regel aus episodischen Gedächtnisaufgaben bestehen (Seron, 2014). In Anlehnung an Merten (2011, 2014) wird in Tabelle 5 ein Überblick über die im deutschen Sprachraum verfügbaren PVT und deren Einsatzhäufigkeit nach Dandachi-FitzGerald et al. (2015) dargestellt. Die *Testbatterie für Forensische Neuropsychologie (TBFN)*: Heubrock & Petermann, 2000) ist mit 41% der am häufigsten verwendete PVT. Der *Word Memory Test (WMT)*: Green, 2003, Revised 2005) liegt mit 36% an zweiter Stelle, der *Rey 15-Item Memory Test (FIT)*: Rey, 1958) mit 31% an dritter Stelle. Die Einsatzhäufigkeit der deutschen Version des *Amsterdam Short-Term Memory Tests (ASTM)*: Schmand, Lindeboom, Merten & Millis, 2005a), also des *Amsterdam Kurzzeit Gedächtnistests (AKGT)*: Schmand et al., 2005b) liegt bei 19%. Im europaweiten Vergleich wird der *Amsterdam Short-Term Memory Test (ASTM)* mit 32.4% in der neuropsychologischen Praxis am häufigsten eingesetzt. Der *Rey 15-Item Memory Test (FIT)* liegt im europäischen Vergleich mit 30.4% an zweiter Stelle und der *Word Memory Test (WMT)* findet zu 24.1% Verwendung (Dandachi-FitzGerald et al., 2015; Dandachi-FitzGerald et al., 2013). Im nordamerikanischen Raum wird der *Test of Memory Malingering (TOMM)*: Tombaugh, 1996) mit 77.8% am häufigsten verwendet. Der *Word Memory Test (WMT)* findet zu 58.2% Verwendung, der *Medical Symptom Validity Test (MSVT)*: Green, 2004) zu 28.2%, der *Dot Counting Test (DCT)*: Lezak, 1983) zu 24.7% und der *Rey 15-Item Memory Test (FIT)* zu 16.1% (Martin et al., 2015).

Gemäß Forschungsliteratur wurden der *Word Memory Test (WMT)* und der *Test of Memory Malingering (TOMM)* bislang am meisten untersucht, am besten validiert und am häufigsten eingesetzt (Fazio, Sanders & Denney, 2015; Greve, Ord, Curtis, Bianchini & Brennan, 2008).

Tabelle 5
Modifizierte Kurzbeschreibung deutschsprachiger PVT nach Merten (2011, 2014) und deren Einsatzhäufigkeit laut Dandachi-FitzGerald et al. (2015).

Name des PVT (Abkürzung; Autoren)	Stimulusmaterial (Itemzahl)	Kurzbeschreibung und Bewertung	Einsatz %
Aggravations- und Simulationstest (AST: Eberl & Wilhelm, 2007)	Ziffern, Wörter, Zeichnungen (70)	Alternativwahlformat mit wechselndem Stimulusmaterial, konsekutive Zielitem-Präsentation, offensichtliche Leichtigkeit, keine unabhängige empirische Validierung	15
Amsterdamer Kurzzeitgedächtnistest (AKGT: Schmand et al., 2005b)	semantisch assoziierte Wörter (30)	Mehrfachwahlformat mit mathematischen Distraktoraufgaben; solide konstruiert, robust gegenüber Coaching, keine Verwendung bei Patienten mit offensichtlichen schweren kognitiven Störungen	19
Testbatterie für Forensischen Neuropsychologie (TBFN: Heubrock & Petermann, 2000)	Objektabbildungen (100)	Unterschiedliche Subtests (Gedächtnis, visuell-figurale und akustische Wahrnehmung), konsekutive Zielitem-Präsentation, offensichtliche Leichtigkeit, keine unabhängige empirische Validierung	41
Coin-in-the-Hand-Test, (Kapur, 1994)	Münzen (10)	Schnell, einfach, kostenfrei, nur zur Erfassung sehr grober Täuschungsversuche, geringe Augenscheinvalidität und Teststärke	4
Computerized Assessment of Response Bias (CARB: Allen, Conder, Green & Cox, 1997)	Zahlenfolgen (111)	Keine deutschsprachige empirische Validierung, kein deutsches Testmanual, anfällig für Coaching	k. A.
Fifteen Item Test (FIT: Rey, 1958)	Zeichen	Präsentation von Buchstaben, Zahlen, Figuren mit anschließender Reproduktion; schnell, einfach, kostenfrei; traditionelles Verfahren, nur als Ergänzung, nicht als alleinige Methode ausreichend	31
Medical Symptom Validity Test (MVST: Green, 2004)	Assoziierbare Wortpaare (20)	Alternativwahlformat; Kurzverfahren basierend auf WMT mit vereinfachten Anforderungen; Profilvergleich mit schweren kognitiven Störungen; empirische Validierung; nur PC-Version, kein deutsches Testmanual	5
Medical Symptom Validity Test (NV-MVST: Green, 2004)	Objektabbildungen	Basierend auf WMT und MSVT; nonverbale Erfassung der Anstrengungsbereitschaft sowie tatsächlicher kognitiver Störungen, demenzielle Normgruppe, bisher eingeschränkte empirische Studienlage, kein deutsches Testmanual	k. A.
Test of Memory Malingering (TOMM: Tombaugh, 1996)	Strichzeichnungen (Objekte, Tiere) (50)	Sehr gute Validierung durch zahlreiche unabhängige Studien; sprachfrei; hohe Augenscheinvalidität; PC- und Papier-Bleistift-Version; kein deutsches Testmanual; möglicherweise verminderte Sensitivität	13
Word Memory Test (WMT: Green, 2003, Revised 2005)	Assoziierbare Wortpaare (40)	Probates Alternativwahlverfahren, breite und ständig überarbeitete Normierung; multilinguale Testversionen, demenzielle Normgruppe und kognitives Leistungsprofil bei Patienten mit kognitiven Störungen; PC- und Papier-Bleistift-Version; kein deutsches Testmanual; hohe Sensitivität bei mögl. verminderte Spezifität	36

Diese Verfahren werden daher auch häufig zur Validierung neuer Verfahren verwendet. Beispielsweise wurde der *WMT* von Bodenburg (2014), Davis und Wall (2014) sowie Donders und Strong (2011) als Außenkriterien zur Ermittlung eingebetteter Parameter verminderter Anstrengungsbereitschaft herangezogen.

2.8.2 Gütekriterien der PVT

Greiffenstein, Greve, Bianchini und Baker (2008) betonen mit dem zunehmenden Einsatz der PVT die Relevanz ihrer Gütekriterien. Die konventionellen psychometrischen Gütekriterien sind jedoch nur von eingeschränkter Bedeutung. Dies liegt darin begründet, dass sich der Streuungsbereich der Ergebnisse meist entweder auf sehr gute Testleistungen (Deckeneffekte) oder eine suboptimale Anstrengungsbereitschaft beschränkt. Demzufolge sind Reliabilitäts-schätzungen für die sehr guten Testleistungen auf Grund der geringen Ergebnisvarianz wenig aussagekräftig. Da die Retest-Reliabilität für suboptimale Anstrengungsbereitschaft stark von der jeweiligen Testsituation abhängt, erscheint deren Aussagekraft ebenfalls als fraglich (Schmand et al., 2005b). Hartman (2002) verfasste Kriterien zur Beurteilung der Güte von PVT, die von Merten (2005) ins Deutsche übersetzt wurden:

1. Sensitivität: Trefferquote eines PVT tatsächlich vorliegende Antwortverzerrungen durch ein auffälliges Testergebnis abzubilden.
2. Spezifität: Trefferquote eines PVT nicht-verminderte Anstrengungsbereitschaft durch ein unauffälliges Testergebnis abzubilden.
3. Resistenz gegenüber Coaching: gezielter Einsatz von Informationen im Vorfeld einer neuropsychologischen Untersuchung zur glaubhaften Präsentation vorgetäuschter kognitiver Beeinträchtigungen.
4. Augenscheinvalidität: Erfassung der mutmaßlichen Zielfunktion mit PVT erscheint dem Untersuchten einleuchtend.
5. Vorliegen reliabler Normen: jeweils für eine gesunde, eine klinische und eine Simulationsreferenzgruppe.
6. Einfachheit des PVT in Anwendung und Auswertung.
7. Einfachheit der Ergebnisinterpretation.
8. Anregungen für aktuelle Forschungsfragen.
9. Hohe Überschneidung des simulierten Funktionsbereichs (z.B. Gedächtnis) und dem kognitiven Funktionsbereich, der durch den PVT erfasst wird.

Anhand dieser Kriterien evaluierte Hartman (2002) acht gängige PVT. Der *Word Memory Test (WMT)* erfüllte als einziges Verfahren alle Kriterien. Der bereits erwähnte *Test of Memory Malingering (TOMM)* und das *Computerized Assessment of Response Bias (CARB: Allen et al., 1997)* erzielten jeweils sieben Kriterien. Der *Rey 15-Item Memory Test (FIT)* und der *Dot Counting Test (DCT)* konnten nur durch die Einfachheit in Anwendung und Auswertung überzeugen (Hartman, 2002; Merten, 2005).

Während die Gütekriterien der Selbstausskunftstests, insbesondere der Versionen des *Minnesota Multiphasic Personality Inventory (MMPI)*, in mehreren Metaanalysen (Green & Rosenfeld, 2011; Hawes & Boccaccini, 2009; Hawes, Boccaccini & Murrie, 2013; Nelson, Hoelzle, Sweet, Arbisi & Demakis, 2010; Rogers, Sewell, Martin & Vitacco, 2003) evaluiert wurden, liegen für PVT bisher nur die beiden Metaanalysen von Sollman und Berry (2011) sowie Vickery et al. (2001) vor.

Vickery et al. (2001) bezogen in ihre Metaanalyse fünf PVT ein, für die zum damaligen Zeitpunkt die meisten empirischen Studien vorlagen. Aus 32 *known-group Studien* mit klinischen Simulationspatienten wurden die Ergebnisse des *Rey 15-Item Memory Test (FIT)*, *Dot Counting Test (DCT)*, *Digit Memory Test (DMT: Hiscock & Hiscock, 1989)*, *Portland Digit Recognition Test (PDRT: Binder, 1993)* und *21-Item Test (Iverson, Franzen & McCracken, 1991)* hinsichtlich der diagnostischen Güte überprüft. Testübergreifend wurde mit $Hedges' d = 1.13$ (95%-CI = 1.04 - 1.22) eine große Effektstärke ermittelt. Die fünf Testverfahren zeigten im Allgemeinen eine moderate durchschnittliche Sensitivität von 56.0% (Spannweite = 22.0% - 83.4%) und eine sehr hohe durchschnittliche Spezifität von 95.7% (Spannweite = 92.5% - 100%). Zehn Jahre nach dem Überblick von Vickery et al. (2001) folgte die Metanalyse von Sollman und Berry (2011). Anhand der Ergebnisse aus 41 *known-group Studien* wurden jeweils die diagnostische Güte des *Word Memory Test (WMT)*, *Test of Memory Malingering (TOMM)*, *Medical Symptom Validity Test (MSVT: Green, 2004)*, *Victoria Symptom Validity Test (VSVT: Slick, Hopp, Strauss & Spellacy, 1996)* sowie des *Letter Memory Test (LMT: Inman et al., 1998)* evaluiert. Analog zu Vickery et al. (2001) wurde für die Verwendung von PVT zur Erfassung negativer Antwortverzerrungen eine große Effektstärke von $Hedges' d = 1.55$ (95%-CI = 1.48 - 1.63) ermittelt. Die Effektstärken der einzelnen Verfahren lagen zwischen $Hedges' d = 2.77$ (*Victoria Symptom Validity Test, VSVT*) und $Hedges' d = .94$ (*Medical Symptom Validity Test, MSVT*). Für den *Word Memory Test (WMT)* wurde eine vergleichbare Effektstärke von $Hedges' d = 1.61$ wie für den *Test of Memory Malingering (TOMM)* und *Letter Memory Test (LMT)* berechnet. Im Durchschnitt

wiesen die PVT eine Spezifität von 90% auf, mit einer Spannweite zwischen 95.5% (*Victoria Symptom Validity Test, VSVT*) und 69.4% (*Word Memory Test, WMT*). Die durchschnittliche Sensitivität lag bei 69% und rangierte zwischen 81.5% (*Victoria Symptom Validity Test, VSVT*) und 65.4% (*Test of Memory Malingering, TOMM*). Für den *Word Memory Test (WMT)* wurden eine Sensitivität von 75.1% und die niedrigste Spezifität von 69.4% ermittelt. Die Gütekriterien der PVT aus den beiden Metaanalysen werden in Tabelle 6 aufgeführt. Die Autoren sehen die geringe Spezifität des *Word Memory Test (WMT)* primär in der Selektion der Studien begründet. Von 64 Referenzen entsprachen nur sechs Studien den Auswahlkriterien, um in die Analysen einbezogen zu werden.

Tabelle 6

Klassifikationsstatistiken der PVT in den Metaanalysen von Vickery et al. (2001) und Sollman & Berry (2011).

	Vickery et al. (2001)				Sollman & Berry (2011)				
	<i>DMT</i>	<i>PDRT</i>	<i>FIT</i>	<i>21-Item-Test</i>	<i>VSVT</i>	<i>WMT</i>	<i>TOMM</i>	<i>LMT</i>	<i>MSVT</i>
Sensitivität %	83.55	44.01	43.27	21.96	81.5	75.1	65.4	70.2	70.0
Spezifität %	95.12	97.25	92.50	100.00	95.5	69.4	93.8	93.0	91.3
Hit Rate %	89.38	71.15	70.45	60.65	88.5	72.3	79.6	81.6	80.7

Aus den bisherigen Überblicksarbeiten schließen Bashem et al. (2014), dass die Spezifität validierter PVT in der Regel mindestens 90% beträgt, während die Sensitivität meist zwischen 50% und 60% variiert (Boone, 2007; Larrabee et al., 2007; Vickery et al., 2001).

2.8.3 Interpretation der PVT

Insgesamt gelten PVT als die am besten standardisierten (Merten, 2009) und zuverlässigsten (Merten, 2005) Verfahren zur Validierung kognitiver Defizite. Von allen Methoden der BV (s.o.) weisen sie die größte Objektivität auf und sind am stärksten von der Person des Gutachters unabhängig. In der Regel sind sie einfach in Anwendung und Auswertung, so dass ihnen eine hohe Praktikabilität zugesprochen werden kann (Merten, 2011). Für Seron (2014) bestehen Nachteile darin, dass die meisten PVT die Anstrengungsbereitschaft ausschließlich mit Hilfe von Aufgaben zum deklarativen bzw. episodischen Gedächtnis erfassen. Patienten beklagen jedoch häufig Beeinträchtigungen in anderen kognitiven Funktionsbereichen wie Aufmerksamkeit, Wahrnehmung, motorische Funktionen und problemlösendes Denken. Somit besteht ein Mangel an Verfahren, die auch diese Funktionsbereiche berücksichtigen. Des Weiteren haben viele PVT auf Grund ihrer Einfachheit eine geringe Sensitivität, da

negative Antwortverzerrungen seitens der Patienten an das Schwierigkeitsniveau angepasst werden, um keine zu offensichtlichen Defizite zu präsentieren (Seron, 2014). Zudem sind PVT vergleichsweise einfach zu durchschauen. Dadurch sind das Testprinzip und die Testintention leicht identifizierbar, insbesondere nach vorheriger gezielter Informationssuche oder vorbereitendes Coaching (Merten, 2014). Darüber hinaus besteht die Gefahr von Fehlklassifikationen. Zum einen durch geringe, kontextabhängig variierende Prävalenzraten, zum anderen besteht bei einem auffälligen PVT das Risiko einer Übergeneralisierung durch die Vergabe einer nicht gerechtfertigten Simulationsdiagnose (Seron, 2014). Dies ist unter anderem darin begründet, dass auffälligen Testergebnissen mehr Bedeutung beigemessen wird. Während ein auffälliger PVT als Hinweis für verminderte Anstrengungsbereitschaft in den diagnostischen Prozess einfließt, können unauffällige Ergebnisse Zweifel an der Anstrengungsbereitschaft nicht ausschließen (Heilbronner et al., 2009).

Laut der DGPPN ist der Einsatz von PVT als einzelner „Mosaikstein“ (Dressing et al., 2011, S. 3, Z. 18) im Rahmen einer umfassenden Begutachtung zu betrachten. PVT sollen in eine neuropsychologische Standarduntersuchung integriert werden, um die Ergebnisse vergleichen zu können. Bei einem PVT unterhalb der Ratewahrscheinlichkeit kann von einer bewussten Antwortmanipulation ausgegangen werden. Ein Wert unterhalb des kritischen Schwellenwertes eines PVT gilt lediglich als Hinweis für geringe Anstrengungsbereitschaft (Rüsseler et al., 2008). Darauf folgt der eigentlich relevante Schritt: die Interpretation der gezeigten Testperformanz im klinischen Gesamtkontext und die Entscheidung über das Vorliegen von Aggravations- oder Simulationstendenzen durch den Gutachter (Dreßing, Widder & Foerster, 2010). Demzufolge wird die Frage, inwieweit die Testperformanz einer Aggravation oder Simulation zuzuordnen ist, von einem Gutachter und nicht von einem PVT beantwortet (Dressing et al., 2011). Dies betonen auch diverse Autoren wie bspw. Merten (2014), Eberl und Wilhelm (2007) sowie Schmand et al. (2005b) und entspricht der Auffassung in einer Umfrage von Dandachi-FitzGerald et al. (2015). Die deutschsprachigen neuropsychologischen Sachverständigen bezogen sich zur Einschätzung der Anstrengungsbereitschaft auf die Beurteilung von Diskrepanzen zwischen Aktenlage, Symptomschilderungen und dem beobachtbaren Verhalten. Zusammenfassend wird von der DGPPN und zahlreichen Autoren (Bigler, 2015; Dressing et al., 2011; Dreßing et al., 2010; Heilbronner et al., 2009; Merten et al., 2010) die Ansicht vertreten, dass PVT stets als Ergänzung einer Plausibilität- und Konsistenzprüfung und nicht als allein gültiges Diagnostikum für Simulation und Aggravation zu sehen sind.

PVT stellen die bisher bedeutsamsten neuropsychologischen Methoden zur Erfassung negativer Antwortverzerrungen dar. Es handelt sich meist um Gedächtnistests, die nach dem Prinzip der verdeckten Leichtigkeit konstruiert sind und auf Wahrscheinlichkeitsannahmen über Zufallsantworten basieren. Ein Ergebnis unterhalb der Ratewahrscheinlichkeit wird bei Vorliegen eines externalen Anreizes als sichere Simulation interpretiert. Insgesamt wird den PVT im diagnostischen Prozess eine ergänzende Funktion zugesprochen. Die relevanten Testgütekriterien sind Sensitivitäts- und Spezifitätsstatistiken. Zur Vermeidung von Falsch-Positiv-Fehlklassifikationen weisen gut validierte PVT eine Spezifität von mindestens 90% auf, was jedoch mit einer geringeren Sensitivität für tatsächlich vorliegende Antwortverzerrungen einhergehen kann.

In den nächsten Abschnitten werden die empirischen Studienergebnisse zu den drei PVT *Word Memory Test (WMT)*, *Amsterdam Kurzzeit Gedächtnistest (AKGT)* und des *Aggravations- und Simulationstest (AST 4.0)*; Eberl & Wilhelm, 2007) aufgeführt, die in der vorliegenden Arbeit verwendet wurden. Der Fokus liegt auf der empirischen Validierung, eine ausführliche Beschreibung der Verfahren ist im Methodenteil Abschnitt 4.2.3 zu finden.

2.8.4 Word Memory Test (WMT)

1994 wurde der *Word Memory Test (WMT)* von Paul Green entwickelt. Mit dem *WMT* kann neben der Anstrengungsbereitschaft auch die Gedächtnisleistung erfasst werden. Die Anstrengungsbereitschaft wird mit den drei Kennwerten *unmittelbares Wiedererkennen*, *verzögertes Wiedererkennen* und dem *Konsistenzmaß* ermittelt. Anhand des Profils tatsächlicher Gedächtnisstörungen (*Genuine Memory Impairment Profile*) kann zusätzlich eine Diskrepanz zwischen einfachen und schwierigen Aufgaben verdeutlicht werden, die typisch für tatsächlich vorliegende Gedächtnisstörungen ist (Green, 2003, Revised 2005). Der Zusammenhang zwischen auffälligen *WMT*-Werten und verminderten Testleistungen in neuropsychologischen Standardverfahren wurde bereits in mehreren Arbeiten ermittelt (Green et al., 2001; Merten, Bossink & Schmand, 2007; Rohling & Demakis, 2010). Die auffälligen *WMT*-Werte korrelierten dabei mit verminderten Testleistungen in den Bereichen Gedächtnis, Sensomotorik, Exekutivfunktionen, Aufmerksamkeit, verbales Verständnis und Wahrnehmungsorganisation (Green et al., 2001).

Während Sollman und Berry (2011) in ihrer Metaanalyse vergleichsweise niedrige Sensitivitäts- und insbesondere Spezifitätswerte für den *WMT* ermittelten, ergaben zahlreiche Untersuchungen zuverlässige Klassifikationen mit dem *WMT*, der als einer der am besten

validiertesten PVT gilt (Armistead-Jehle & Buican, 2013; Donders & Strong, 2013; Fazio et al., 2015; Merten, 2011; Merten et al., 2004). In den Analogstudien wurden sehr zuverlässige Klassifikationsstatistiken ermittelt. Bei Tan, Slick, Strauss und Hultsch (2002) mit 56 klinisch unauffälligen Probanden zeigte der *WMT* im Vergleich zum *Test of Memory Malingering (TOMM)* und zum *Victoria Symptom Validity Test (VSVT)* die höchste Klassifikationsgenauigkeit mit 96.3%. Die Vorhersage der Gruppenzugehörigkeit (KG Kontrollgruppe vs. ES experimentelle Simulanten) erfolgte anhand einer logistischen Regressionsanalyse und ergab für den *WMT* eine Sensitivität von 92.6% und eine Spezifität von 100%. Auch Brockhaus und Merten (2004) fanden in einer Analogstudie mit 127 klinisch unauffälligen Probanden eine hervorragende Klassifikationsgenauigkeit von 100% für die deutschsprachige Version des *WMT*. Ergänzend untersuchten die Autoren eine klinische Stichprobe mit 32 Patienten mit unterdurchschnittlicher intellektueller Leistungsfähigkeit. Mit Ausnahme eines Patienten lagen die *WMT*-Werte sämtlicher Patienten oberhalb des Schwellenwerts. Die diagnostische Güte des *WMT* wurde durch die hohe Klassifikationsgenauigkeit von 97% bestätigt. Eine weitere Analogstudie veröffentlichten Martins und Martins (2010) mit einer portugiesischen Version des Verfahrens. Neben 49 klinisch unauffälligen Probanden wurden 21 Patienten mit objektivierbaren leichten Gedächtnisproblemen bei erhaltener Alltagskompetenz (*Mild Cognitive Impairment*) in die Studie einbezogen. Aus der Vorhersage der experimentellen Gruppenzugehörigkeit der gesunden Probanden (KG Kontrollgruppe vs. ES experimentelle Simulanten) ermittelten die Autoren eine Sensitivität von 95% und eine Spezifität von 100%. Allerdings erzielten 67% der Patienten mit Gedächtnisproblemen auffällige Testwerte, so dass sich eine hohe Rate von Falsch-Positiv-Fehlklassifikationen ergab. Eine Verbesserung der Sensitivität und Spezifität auf jeweils 95% wurde durch die Modifikation des Schwellenwertes von 82.5% auf 60% in den beiden Rekognitionsdurchgängen oder durch Berücksichtigung der freien Abrufleistung erzielt. Wie Merten, Bossink, et al. (2007) nahmen auch Martins und Martins (2010) an, dass durch Herabsetzen der Schwellenwerte und die Berücksichtigung ergänzender Gedächtnisparameter die diagnostische Güte des *WMT* verbessert werden kann. Insgesamt ergaben sich aus den Analogstudien für die drei *WMT*-Kennwerte jeweils hohe Klassifikationsgenauigkeiten zwischen 93% und 100%.

Die zuverlässigen Klassifikationsstatistiken der Analogstudien zeigten sich auch mehrheitlich in einer hohen diagnostischen Güte des *WMT* in *known-group Studien*. Gervais et al. (2004) verglichen den *WMT* mit dem *Test of Memory Malingering (TOMM)* und dem *Computerized*

Assessment of Response Bias (CARB) in einer Patientenstichprobe mit objektivierbarem Anreiz zur Symptomverstärkung ($N = 519$). Insgesamt zeigten 35% der Patienten ein auffälliges Ergebnis in mindestens einem PVT. 32% der Patienten erzielten auffällige Werte im *WMT*, 11% im *TOMM* und 17% im *CARB*. Mit Ausnahme eines Patienten zeigten alle Patienten, die den *TOMM* nicht bestanden hatten, auch auffällige *WMT*-Werte. Darüber hinaus bestanden 109 Personen (21%) den *TOMM*, während sie sowohl im *WMT* als auch im *CARB* auffällige Werte präsentierten. Auf Grund dessen schlossen die Autoren auf eine höhere Sensitivität des *WMT* im Vergleich zu den beiden anderen PVT. Weiter sahen die Autoren ihre Annahme durch die Berücksichtigung des kognitiven Status bestätigt, da Patienten mit auffälligem *WMT* einen besseren kognitiven Status zeigten, als Patienten, die keinen der drei PVT bestanden. In einer weiteren *known-group Studie* mit 473 neurologischen bzw. psychosomatischen Patienten und externalem Anreiz zur Symptomverstärkung konnten Greiffenstein et al. (2008) die höhere Sensitivität des *WMT* im Vergleich zum *TOMM* replizieren. In einer asymmetrischen Konkordanzanalyse zeigten 21.4% der Patienten auffällige Werte in mindestens einem der drei *WMT*-Werte und dem *TOMM*. Zudem erzielten 27.3% der Patienten auffällige Werte in mindestens einem der *WMT*-Kennwerte bei gleichzeitig unauffälligem *TOMM*. Umgekehrt zeigten nur 1% einen auffälligen *TOMM*-Wert bei gleichzeitig unauffälligen *WMT*-Kennwerten. Obwohl in einer symmetrischen Konkordanzanalyse mit jeweils drei *WMT*- und *TOMM*-Kennwerten die Sensitivität des *TOMM* verbessert wurde, erzielten weiterhin mehr Patienten auffällige Werte im *WMT*. Dieses Ergebnis führten die Autoren auf die höhere Sensitivität des *WMT* zurück.

Die Gütekriterien des *WMT* und *TOMM* wurden auch mit einer forensischen Stichprobe validiert. Fazio et al. (2015) unterteilten die Stichprobe entsprechend der Ergebnisse in zwei Gruppen (verminderte Anstrengungsbereitschaft vs. unauffällige Anstrengungsbereitschaft). Die Klassifikationsstatistiken des *WMT* wiesen zwar eine hervorragende Spezifität von 94.7% auf, während die Sensitivität von 56.1% kaum aussagekräftig war. Die Spezifität des *TOMM* lag dagegen bei 86.8%, die Sensitivität bei 68.2%. Durch eine Kombination des *WMT* und des *TOMM* wurden eine Sensitivität von 68.3% und eine Spezifität von 92.1% erreicht. Bei einer Prävalenzrate von 50% für das Vorliegen negativer Antwortverzerrungen zeigte sich der *WMT* dem *TOMM* hinsichtlich einer sehr hohen positiven Vorhersagevalidität von 91.4% als diagnostisch überlegen. Dagegen wies der *TOMM* eine bessere negative Vorhersagevalidität (73.2%) als der *WMT* (68.3%) auf.

Darüber hinaus wurde die diagnostische Güte des *WMT* in diversen *known-group Studien* mit Schädelhirntrauma-Patienten überprüft. In den Validierungsstudien von Green (2003) und Green et al. (2001) erzielten Patienten mit moderaten bis schweren Schädelhirntraumata im *WMT* eine durchschnittliche Rekognitionsleistung von 95%. Die Rekognitionsleistungen von gesunden Kontrollpersonen rangierten zwischen 97.8% und 99.5%. Zudem zeigte sich bei Gutachtenpatienten in mehreren Studien ein inverser Zusammenhang zwischen der Schwere des Schädelhirntraumas und der Präsentation auffälliger Werte im *WMT* (Flaro, Green & Robertson, 2007; Green, Flaro & Courtney, 2009; Green, Iverson & Allen, 1999). Bei Flaro et al. (2007) erzielten 21% der Patienten mit moderaten bis schweren Schädelhirntraumata auffällige *WMT*-Werte. Die Anzahl der Patienten mit leichtem Schädelhirntrauma, die den *WMT* nicht bestand, war mit 40% fast doppelt so hoch. Diesen Zusammenhang konnten die Autoren auch anhand bildgebender Verfahren verdeutlichen. Patienten mit auffälligen CT- oder MRT-Scans präsentierten häufiger bessere Leistungen im *WMT* als Patienten mit unauffälligen Scans.

Andererseits deutet insbesondere die Metaanalyse von Sollman und Berry (2011) auf eine unzureichende Spezifität und damit eine erhöhte Rate an Falsch-Positiv-Fehlklassifikationen des *WMT*. Dies veranschaulicht auch die Studie von Merten, Bossink, et al. (2007) mit 48 neurologischen Patienten ohne objektivierbaren Anreiz für negative Antwortverzerrungen. Von den Patienten mit klinisch bedeutsamen Gedächtnisdefiziten erzielten 50% bis 58% auffällige Werte im *WMT*, wiederum fast alle Patienten ohne kognitive Störung (92% bis 100%) bestanden den *WMT*. Zur Differenzierung zwischen verminderter Anstrengungsbereitschaft und tatsächlichen kognitiven Beeinträchtigungen entwickelte Green (2005) ein Profil tatsächlicher Gedächtnisstörungen (*Genuine Memory Impairment Profile*) für den *WMT*. Während sich bei Martins und Martins (2010) sowie Fazio et al. (2015) die Spezifität durch das Gedächtnisprofil verbessert ließ, konnten Sherer et al. (2015) diese Ergebnisse nicht replizieren. Auch bei Gorissen, Sanz und Schmand (2005) wurden mit dem *WMT* in einer gemischten psychiatrischen Stichprobe ohne Simulationsverdacht zahlreiche Falsch-Positive ermittelt. 72% der psychotischen Patienten zeigten auffällige Werte im *WMT*, ebenso 25% der nicht-psychotischen, meist affektiv erkrankten Patienten. Die Fehlklassifikationen von Patienten mit schweren (kognitiven) Störungen, ebenso wie die Ergebnisse bildgebender Verfahren (Allen et al., 2007; Larsen et al., 2010), stellen die Annahme infrage (s.o.), inwiefern bei der Bearbeitung des *WMT* tatsächlich nur wenig kognitive Ressourcen beansprucht werden.

Der *WMT* wurde in mehreren Analogstudien mit hervorragenden Klassifikationsstatistiken validiert. In *known-group Studien* konnte die experimentell ermittelte hohe Sensitivität bestätigt werden, die Spezifität des *WMT* wurde jedoch nicht immer repliziert. Zudem deuten bildgebende Verfahren und schwer beeinträchtigte Patientenstichproben auf eine stärkere kognitive Beanspruchung durch den *WMT* hin als auf Grund der Studien mit Schädelhirntrauma-Patienten angenommen wurde. Dennoch gilt der *WMT* als gut etablierter PVT und wird häufig in Forschung und Praxis eingesetzt.

2.8.5 Amsterdamer Kurzzeitgedächtnistest (AKGT)

Der solide konstruierte *Amsterdam Short-Term Memory Tests (ASTM)*: Schmand et al., 2005a) ist der in Europa am häufigsten eingesetzte PVT (Merten, 2011). Im nordamerikanischen Sprachraum ist er dagegen weitgehend unbekannt (Dandachi-FitzGerald et al., 2013).

Im Testmanual der deutschsprachigen Version des *Amsterdam Kurzzeit Gedächtnistests (AKGT)*: Schmand et al., 2005b) verweisen die Autoren auf die erste Analogstudie zur Validierung der niederländischen Version mit dem *Rey 15-Item Memory Test (FIT)* und dem *Dot Counting Test (DCT)*. Für den *Amsterdam Short-Term Memory Test (ASTM)* wurden eine Sensitivität von 72% und eine Spezifität von 100% ermittelt (Schagen et al., 1997). In einer weiteren Analogstudie von Schagen et al. (1997) bearbeiteten neben experimentellen Simulanten (ES) und einer Kontrollgruppe (KG) auch Schädelhirntrauma-Patienten ohne externalen Anreiz den *ASTM*. Während sich in den Deckeneffekten der KG ($M = 89.7$, $SD = 0.7$) und Schädelhirntrauma-Patienten ($M = 89.4$, $SD = 1.1$) keine statistisch bedeutsamen Unterschiede zeigten, erzielten die ES ($M = 81.0$, $SD = 3.0$) signifikant schlechtere Ergebnisse. Die Klassifikationsgenauigkeit betrug 100%, die interne Konsistenz Cronbach's $\alpha = .87$. Bolan et al. (2002) untersuchten die Vorhersagegenauigkeit des englischsprachigen *ASTM* (Schmand, de Sterke & Lindeboom, 1999) mit 32 gesunden Probanden. Darüber hinaus wurden der *Test of Memory Malingering (TOMM)* und der *Digit Recognition Test (DRT)*, der von den Autoren für diese Studie entwickelt wurde, zur Erfassung der Anstrengungsbereitschaft verwendet. Im *ASTM* zeigten die ES ($M = 58.6$, $SD = 9.3$) eine signifikant schlechtere Testperformanz als die KG ($M = 89.1$, $SD = 1.1$). Die Klassifikationsgenauigkeit des *ASTM* und *TOMM* lag jeweils bei 100%, da alle Probanden gemäß ihrer Gruppenzugehörigkeit korrekt klassifiziert wurden. Zudem schätzten 72% der Probanden den *ASTM* als anspruchsvollen Gedächtnistest für Patienten mit kognitiven Beeinträchtigungen ein. Daraus schlossen die Autoren auf eine hohe Augenscheinvalidität des

Verfahrens und eine erfolgreiche Operationalisierung des Prinzips der verdeckten Leichtigkeit. In einer weiteren Analogstudie von Jelacic, Merckelbach, Candel und Geraerts (2007) wurde die diagnostischen Güte des *ASTM* im Vergleich zum *Structured Inventory of Malingered Symptomatology (SIMS)* untersucht. Die Stichprobe setzte sich aus jeweils 30 KG, ES und speziell instruierten experimentellen Simulanten (CS) zusammen, die ein Coaching mit Zusatzinformationen über die kognitiven Symptome sowie Warnungen vor zu offensichtlichen Symptomdarstellungen erhielten. Die beiden Simulationsgruppen ES und CS erzielten im *ASTM* schlechtere Werte als die KG ($M = 88.1, SD = 1.6$) und die ES ($M = 68.4, SD = 14.0$) präsentierten auffälliger Werte als die CS ($M = 77.3, SD = 9.7$). Anhand des *ASTM* konnten jeweils 90% der KG und ES sowie 70% der CS Simulanten korrekt klassifiziert werden. Die Ergebnisse zeigten eine Rate von 10% Falsch-Positiv-Klassifikationen durch den *ASTM*, während mit dem *SIMS* alle KG zu 100% korrekt klassifiziert wurden.

Darüber hinaus wurde auch die Klassifikationsgüte des deutschsprachigen *Amsterdam Kurzzeit Gedächtnistests (AKGT)* in mehreren Studien empirisch überprüft. Merten et al. (2004) verglichen in einer Analogstudie die Klassifikationsgüte des *AKGT*, *Word Memory Test (WMT)* und *Word Completion Memory Test (WCMT)*: Hilsabeck et al., 2001). Die Rate korrekter Klassifikationen lag insgesamt zwischen 95% und 100%. Für den *WMT* wurde eine Sensitivität und Spezifität von 100% ermittelt. Die Sensitivität des *AKGT* lag ebenfalls bei 100%, die Spezifität bei 90%. Auch Merten, Green, Henry, Blaskewitz und Brockhaus (2005) führten eine Analogstudie mit jeweils 18 KG, ES und CS durch. Die Probanden bearbeiteten den *AKGT*, *Rey 15-Item Memory Test (FIT)* und *Medical Symptom Validity Test (MSVT)*. Die Studie umfasste auch die neuropsychologischen Standardverfahren *Trail-Making-Test Teil A* und *Trail-Making-Test Teil B (TMT-A, TMT-B)*: Reitan, 1992), *Rey Complex Figure Test (CFT)*: Osterrieth, 1944), *Digit Span* zur Erfassung der Zahlenspanne der Wechsler-Skalen (Wechsler, 1981), woraus die beiden eingebetteten Parameter, das Verhältnis *TMT-B/TMT-A* sowie *Reliable Digit Span (RDS)*, berechnet wurden. Es zeigten sich signifikant schlechtere Leistungen der CS in allen Leistungsparametern mit Ausnahme des Verhältnisses *TMT-B/TMT-A*. Die Klassifikationsgenauigkeit der PVT rangierte zwischen 56% (*FIT*) und 100% (*AKGT, MVST*). Die Klassifikationsgenauigkeit der eingebetteten Parameter lag bei 50% (Verhältnis *TMT-B/TMT-A*) und 86% (*RDS*). In dieser Studie zeigte sich somit eine höhere diagnostische Güte der PVT verglichen mit eingebetteten Parametern.

Auch Blaskewitz und Merten (2006) untersuchten in einer Analogstudie mit 24 klinisch unauffälligen Probanden die Klassifikationsgenauigkeit des *Amsterdam Kurzzeit*

Gedächtnistests (AKGT) im Vergleich zum *Medical Symptom Validity Test (MSVT)* und *One-in-Five Test* (Gubbay, o.J.) sowie den eingebetteten Parametern *Reliable Digit Span (RDS)*, Verhältnis *TMT-B/TMT-A*, Doppel- und Buchstabenfehler des *Aufmerksamkeits-Belastungs-Tests d2* (Brickenkamp, 1994), des *Judgement of Line Orientation Tests* (Benton, Hamsher, Varney & Spreen, 1983) und des *Rey-Osterrieth Complex Figure Tests (RCFT)*: Meyers & Meyers, 1995). Darüber hinaus bestand das Studiendesign aus zwei Messzeitpunkten, so dass die Retest-Reliabilitäten der Verfahren ermittelt werden konnten. In der ersten Erhebung zeigten sich signifikante Gruppenunterschiede in den PVT (*AKGT* $d = -3.59$, *One-in-Five Test* $d = -1.62$, *MSVT* $d = -2.47$) und im *RCFT* beim Abzeichnen (*RCFT-Copy* $d = -1.94$) und beim sofortigen Abrufen der komplexen *Reyfigur* (*RCFT-IR* $d = -2.83$). Für den *AKGT* wurde in dieser Studie die höchste diagnostische Güte ermittelt. In der ersten Erhebung wies der *AKGT* ebenso wie der *MSVT* eine Sensitivität von 92% und Spezifität von 100% auf. In der zweiten Erhebung lag die Klassifikationsgenauigkeit bei 100%. Die Retest-Reliabilität des *AKGT* war zufriedenstellend, der Phi-Koeffizient betrug $\phi = .92$, der Rangkorrelationskoeffizient nach Spearman $r_{tt} = .88$.

Insgesamt basiert die Mehrzahl der hervorragenden Klassifikationsstatistiken des *Amsterdam Short-Term Memory Tests (ASTM)* bzw. *Amsterdam Kurzzeit Gedächtnistests (AKGT)* auf Analogstudien. Der *ASTM* wurde jedoch auch in einigen klinischen Stichproben validiert, wobei die Aussagekraft durch fehlende eindeutige Gruppenzuteilungen, im Sinne eines *known-group Designs*, zum Teil eingeschränkt wird. Beispielsweise verglichen Dandachi-FitzGerald et al. (2011) die Testperformanz im *ASTM* von 183 psychiatrischen Patienten ohne offenkundigen externalen Anreiz zur Symptomproduktion mit den Symptombildungen im *Structured Inventory of Malingered Symptomatology (SIMS)*. 34% der Patienten erzielten in einem oder beiden Verfahren auffällige Werte. Zwischen dem *ASTM* und *SIMS* wurde ein statistisch bedeutsamer negativer Zusammenhang ermittelt. Eine schlechte Testperformanz im *ASTM* korrelierte mit hohen *SIMS*-Werten ($r = -.22$). In einer weiteren Studie untersuchten Dandachi-FitzGerald, van Twillert, van de Sande, van Os und Ponds (2016) mit den beiden Verfahren eine neurologisch-psychiatrische Stichprobe mit $N = 469$ Patienten. 43% der Patienten der Gesamtstichprobe erhielten auf Grund ihrer Erkrankung oder Behinderung Leistungszuwendungen, 17% befanden sich in einem Begutachtungsverfahren. Die Beschwerdvalidität wurde von 276 Patienten der Gesamtstichprobe mit dem *ASTM* und *SIMS* erfasst. 11% erzielten in beiden Verfahren auffällige Werte. Anhand einer binär-logistischen Regressionsanalyse wurde die Gruppenzugehörigkeit entsprechend der *ASTM*-

Testperformanz aus der verbalen Gedächtnisleistung in einem neuropsychologischen Standardverfahren, den Selbstauskünften im *SIMS* und des Vorliegens eines Anreizes vorhergesagt. Mit diesem Regressionsmodell konnte 29% (Nagelkerkes R^2) der Varianz der verminderten Testperformanz aufgeklärt werden. 78% der Patienten konnten mit dem Modell korrekt vorhergesagt werden.

In einer Untersuchung von van Hout, Schmand, Wekking, Hageman und Deelman (2003) mit 145 Patienten mit chronisch toxischen Enzephalopathien zeigten 43% der Patienten auffällige Werte im *ASTM*, 21% erzielten auffällige Ergebnisse im *Test of Memory Malinger* (*TOMM*). Bei einer Spezifität von 99% für beide Verfahren zeigte der *ASTM* eine höhere Sensitivität in diesem klinischen Kontext. Allerdings können Falsch-Positiv-Klassifikationen durch den *ASTM* in dieser Stichprobe nicht ausgeschlossen werden. Des Weiteren aggregierten Schmand et al. (2005b) die Daten von 470 neurologischen Patienten aus vier verschiedenen Untersuchungen. Neben dem *AKGT* wurde auch der *WMT* zur BV eingesetzt. In der klinischen Stichprobe zeigten 42.9% der Patienten auffällige *AKGT*-Werte und 41.6% auffällige Werte im *WMT*. Die beiden Verfahren erzielten zu 79% übereinstimmende Ergebnisse ($\kappa = .57$; $p < .001$). Der Korrelationskoeffizient zwischen dem *AKGT* und gemittelten *WMT*-Werten lag bei $r = .71$ und die Autoren schlossen auf eine vergleichbare Vorhersagevalidität des *AKGT* und *WMT*.

2.8.6 Aggravations- und Simulations-Test (AST)

Der computergestützte *Aggravations- und Simulations-Test (AST)* von Eberl und Wilhelm (2007) basiert auf 70 dichotomen Rekognitionsaufgaben mit verbalem, figuralem und numerischem Stimulusmaterial. Durch einen Wechsel der Modalität, im Sinne einer Änderung der Darstellungsform der Rekognitionsaufgaben im Vergleich zur Reizpräsentation der Stimuli, operationalisierten die Autoren das Prinzip der verdeckten Leichtigkeit.

In einer Analogstudie mit einer neurologischen Patientenstichprobe ($n = 39$) und gesunden Probanden ($n = 83$) ermittelten die Autoren des *AST* einen optimalen Schwellenwert von 90%. Bei 90% richtiger Antworten (*AST*) lagen die Klassifikationsstatistiken Sensitivität und Spezifität jeweils bei 100%. Die Testperformanz der beiden experimentellen Gruppen (experimentellen Simulanten ES, Kontrollgruppe KG) zeigte dabei signifikante Unterschiede. Die ES erzielten schlechtere Testergebnisse als die neurologischen Patienten, die sich wiederum nicht in der Anzahl richtiger Antworten von der KG unterschieden. Zudem erzielten die ES langsamere Reaktionszeiten als die KG und neurologischen Patienten. Allerdings geben die Autoren im Testmanual keine Angaben zu den Maßen der zentralen

Tendenz oder Effektstärken. Weiter beschreiben die Autoren größere Standardabweichungen der Reaktionszeiten der Patientengruppe im Vergleich zu den beiden experimentellen Gruppen.

Bislang wurde der *AST* lediglich in einer unabhängigen Studie von Merten et al. (2010) untersucht. In dieser *known-group Studie* ($N = 398$) wurden neben 223 neurologisch-psychiatrischen Gutachtenpatienten auch 175 Patienten ohne objektivierbaren Anreiz eingeschlossen. 48% der Patienten zeigten mindestens einen auffälligen Wert in einem der Verfahren *AST*, *Word Memory Test (WMT)*, *Strukturierten Fragebogen Simulierter Symptome (SFSS)*, *Rey 15-Item Memory Test (FIT)* oder *Testbatterie für Forensische Neuropsychologie (TBFN)*. Einen auffälligen *WMT*-Wert zeigten 45.1% der Patienten, davon litten wiederum 43.7% an einem Schädelhirntrauma und 46.6% an einer psychischen Störung. Im *AST* erzielten 26.8% der Patienten ein auffälliges Ergebnis (27.3% Schädelhirntrauma-Patienten, 26.1% psychiatrische Patienten). Insgesamt 11 von 25 Patienten präsentierten sowohl im *AST* als auch im *WMT* oder *SFSS* auffällige Werte (44% Richtig-Positive). Kein Patient mit unauffälligen Werten im *WMT* und *SFSS* erzeugte einen auffälligen *AST*-Wert (0% Falsch-Positive). Dagegen wiesen 14 bzw. 6 Patienten einen unauffälligen *AST*-Wert auf, während sie im *WMT* oder *SFSS* bzw. in beiden Verfahren unterhalb der jeweiligen Schwellenwerte lagen (56% bis 24% Falsch-Negative). Der Kappa-Koeffizient als Maß der Interrater-Reliabilität lag bei $\kappa = .23$.

Der *ASTM* bzw. *AKGT* wurde in zahlreichen unabhängigen Analogstudien mit hervorragenden Klassifikationsstatistiken zwischen 90% und 100% validiert. Die Anzahl an *known-group Studien* ist allerdings limitiert. Dennoch zählt der *AKGT*, ebenso wie der *WMT*, zu den elaborierten deutschsprachigen PVT. Beide Verfahren werden auch als Außenkriterium zur Validierung neuer Verfahren herangezogen. In der Validierungsstudie der Testautoren des *AST* wurde eine Klassifikationsgenauigkeit von 100% ermittelt, die in der *known-group Studie* von Merten et al. (2010) nicht repliziert werden konnte.

2.9 Analyse neuropsychologischer Testprofile

Die neuropsychologische Diagnostik kognitiver Defizite basiert auf der Objektivierung der Defizite in psychometrischen Leistungstests. Je nach Fragestellung, Symptomatik und Pathologie werden in neuropsychologischen Untersuchungen entsprechende Testverfahren zur Ermittlung kognitiver Leistungsprofile herangezogen. Dabei werden neben

Gedächtnisfunktionen u.a. auch das Denk- und Sprachvermögen, die Aufmerksamkeitsfunktionen und die Psychomotorik erfasst. In den letzten Jahren wurden nicht nur PVT für die Diagnostik der BV entwickelt, sondern auch weitere Methoden in etablierten neuropsychologischen Standardverfahren, im Sinne eingebetteter Parameter. Unter eingebetteten Parametern verstehen Merten et al. (2013) empirisch ermittelte Grenzwerte oder Indizes, die sich aus neuropsychologischen Standardverfahren ableiten lassen und ebenfalls als Indikatoren der Beschwerdvalidität dienen können. Zahlreiche Autoren (Denning, 2012; Fiene et al., 2015; Meyers & Volbrecht, 2003; Miele, Gunner, Lynch & McCaffrey, 2011; Slick et al., 1999) sehen die Verwendung neuropsychologischer Standardverfahren als das methodische Vorgehen, mit dem negative Antwortverzerrungen am effektivsten und laut Meyers, Galinsky und Volbrecht (1999) wahrscheinlich am validesten ermittelt werden können. Auch die neuropsychologischen Dachverbände NAN (Bush et al., 2005) und AACN (Heilbronner et al., 2009) empfehlen neben separaten, spezifischen PVT auch die Verwendung eingebetteter Parameter. Zum einen erscheint es weit schwieriger, kontinuierlich in mehreren neuropsychologischen Standardverfahren kognitive Beeinträchtigungen vorzutäuschen als in einem einzelnen PVT. Zum anderen ist das Vortäuschen eines kognitiven Leistungsprofils, das einer spezifischen kognitiven Störung zugrunde liegt, komplexer und anspruchsvoller als das Vortäuschen unspezifischer Defizite in einem PVT (Slick et al., 1999). Eine detaillierte Erfassung kognitiver Funktionsbereiche anhand mehrerer Verfahren erschwert somit das Vortäuschen kognitiver Defizite (Merten, 2002). Diese Annahme wird durch Untersuchungen zur intraindividuellen Variabilität der Testperformanz, wie beispielsweise von Strauss et al. (1999), untermauert. Eingebettete Parameter ermöglichen eine Zeitersparnis, da gleichzeitig das kognitive Funktionsniveau und die Anstrengungsbereitschaft erfasst werden. Zudem kann die Anstrengungsbereitschaft auch retrospektiv nach Ende der Testdurchführung untersucht werden (Larrabee, 2007). Darüber hinaus sind eingebettete Parameter laut Bashem et al. (2014) weniger leicht zu durchschauen und weisen dadurch auch eine größere Robustheit gegenüber Coaching auf. Diverse Autoren wie Weinborn, Woods, Nulsen und Leighton (2012), Suhr und Gunstad (2007), Blaskewitz, Gorny, Merten und Kathmann (2007), Merten et al. (2005), Bauer und McCaffrey (2006) sowie Erdal (2004) sehen durch Coaching eine Gefährdung der Testvalidität und betonen die Notwendigkeit einer robusten BV gegenüber Coaching. Als Nachteil eingebetteter Parameter gilt die Konfundierung der Anstrengungsbereitschaft mit tatsächlichen kognitiven Beeinträchtigungen (Bashem et al., 2014). Während PVT zur Beurteilung der Anstrengungsbereitschaft weitgehend unabhängig

des tatsächlichen kognitiven Status einer Person konstruiert werden, dienen neuropsychologische Standardverfahren primär zur Erfassung der kognitiven Leistungsfähigkeit. Demzufolge erscheint es plausibel, dass das kognitive Funktionsniveau einer Person stärker in die Berechnung der eingebetteten Parameter miteinfließt als bei unabhängigen PVT. Insgesamt nimmt die Verwendung eingebetteter Parameter nicht nur in der Forschungsliteratur, sondern auch in der klinischen Praxis zu. Laut einer Umfrage von Martin et al. (2015) werden eingebettete Parameter inzwischen deutlich häufiger verwendet als noch in einer Umfrage acht Jahre zuvor (Sharland & Gfeller, 2007).

Zu den etablierten eingebetteten Parametern zählen insbesondere die *Reliable Digit Span (RDS)* von Greiffenstein et al. (1994) und das Verhältnis der Bearbeitungszeiten im *TMT-B* zu *TMT-A* von Ruffolo, Guilmette und Willis (2000). Die *Reliable Digit Span (RDS)* wird aus der Anzahl korrekt wiedergegebener Ziffernfolgen des Subtests Zahlennachsprechen der Wechsler-Skalen ermittelt. In mehreren Studien wurde ein kritischer Schwellenwert \leq sechs Rohwertpunkten verwendet (Boone, 2013; Greiffenstein et al., 1994). Aus 24 Studienergebnissen ermittelten Jasinski, Berry, Shandera und Clark (2011) in einer Metaanalyse bei einem Schwellenwert von 7.1 eine Klassifikationsgenauigkeit der *RDS* von 75.7% (Sensitivität 63.3%, Spezifität 86.1%). Das Verhältnis *TMT-B / TMT-A* wird aus den Bearbeitungszeiten der beiden Testverfahren berechnet. Darüber hinaus wurden auch die Gedächtnisparameter des *California Verbal Learning Test (CVLT)*: Delis, Kramer, Kaplan & Ober, 1987, 2000) in einigen Studien (Donders & Strong, 2011; Whiteside et al., 2015; Wolfe et al., 2010) und die Aufmerksamkeitsparameter des Subtests *Alertness* der *Testatterie zur Aufmerksamkeitsprüfung (TAP)*: Zimmermann & Fimm, 2002) von Bodenburg (2014) als Indikatoren vermindelter Anstrengungsbereitschaft überprüft. Im Folgenden wird der empirische Forschungsstand zu den eingebetteten Leistungsparametern des *CVLT* und der *TAP Alertness* dargestellt. Die ausführlichen Beschreibungen der beiden Verfahren sind im Methodenteil Abschnitt 4.2.2 zu finden.

2.9.1 Eingebettete Parameter des California Verbal Learning Test (CVLT)

Während das neuropsychologische Verständnis von Gedächtnisstörungen lange durch die Annahme eines einheitlichen Amnesiekonzepts geprägt war, unterscheidet man heute qualitativ verschiedene Gedächtnisstörungen im Rahmen diverser Krankheitsbilder (Niemann et al., 2008). Gedächtnisstörungen können ätiologisch im Rahmen eines allgemeinen intellektuellen Leistungsverlusts, Demenzen, fokalen Hirnschäden, funktionellen oder dissoziativen Amnesien sowie diversen internistischen und psychiatrischen Erkrankungen

begründet sein (Thöne-Otto et al., 2010). Es haben sich Differenzierungen in Kurzzeit- und Langzeitgedächtnis, in anterogrades (Neugeächtnis), retrogrades (Altgedächtnis) und prospektives Gedächtnis etabliert. Inhaltlich wird das Langzeitgedächtnis entsprechend des Abrufmodus in deklarative (semantisches und episodisches Gedächtnis) und nondeklarative Gedächtnisstrukturen (Priming, prozedurales Gedächtnis, Konditionierung, nichtassoziatives Lernen) unterteilt. Zudem gliedert man den Gedächtnisprozess in zeitlicher Hinsicht in die Phasen Enkodierung, Konsolidierung/Speicherung und Abruf der Informationen (Calabrese & Markowitsch, 2003; Thöne-Otto et al., 2010). Die zugrunde liegenden zerebralen Strukturen können nach Thöne-Otto et al. (2010) wie folgt zusammengefasst werden: Die Enkodierung semantischer und episodischer Inhalte des Langzeitgedächtnisses erfolgt primär über frontale und parietale Kortexareale. Bei der Konsolidierung und dem Transfer in das Langzeitgedächtnis ist v.a. das limbischen System beteiligt (Thöne-Otto & Markowitsch, 2004). Die Informationsspeicherung und weitere Konsolidierung involviert vorrangig neokortikale Strukturen mit einer stärkeren Aktivierung der linken Hemisphäre für semantische Informationen und der rechten Hemisphäre für episodische Informationen (Markowitsch, 2002). Der Abruf semantischer Informationen erfolgt über die Regionen des linken Frontal- und Temporallappens, der Abruf autobiografischer Informationen über das analoge frontotemporale System der rechten Hemisphäre (Thöne-Otto et al., 2010). Die neurobiologischen Korrelate mnestischer Störungen werden aus den Funktionseinbußen klinischer Stichproben geschlossen. Darüber hinaus bilden insbesondere neuropsychologische Testverfahren, wie der *California Verbal Learning Test (CVLT)*, die Grundlage der Diagnose von Gedächtnisstörungen. Der *CVLT* besteht aus einer Liste mit 16 zu memorierenden Wörtern und verschiedenen Abruf- und Rekognitionsdurchgängen. Laut einer Umfrage nordamerikanischer Neuropsychologen gehört die zweite Version des *CVLT*, der *CVLT-II* (Delis et al., 2000) nach den *Wechsler Gedächtnisskalen* (Wechsler, 1987) zu den am häufigsten angewandten Gedächtnistests (Rabin, Paolillo & Barr, 2016). Demzufolge erscheint es naheliegend, den *CVLT* zur Diagnostik negativer Antworttendenzen heranzuziehen. Das erfolgreiche Vortäuschen kognitiver Beeinträchtigungen im *CVLT* gilt auf Grund der Komplexität und Vielzahl an Leistungsparametern als unwahrscheinlich (Sweet et al., 2000). Aus diversen Studien mit unterschiedlichen statistischen und methodischen Vorgehensweisen wurden verschiedene eingebettete Parameter des *CVLT* ermittelt.

Trueblood und Schmidt (1993) waren die ersten, die aus den Rohwerten der ersten Version des *CVLT* (Delis et al., 1987), nachfolgend als *CVLT-I* bezeichnet, eingebettete Parameter

ableiteten. Von 106 Schädelhirntrauma-Patienten der Gesamtstichprobe erfüllten acht Patienten das Slick-Kriterium einer sicheren SND. Im Vergleich zu ihren gematchten Kontrollpatienten im abschließenden Rekognitionsdurchgang des *CVLT-I* erzielten sie eine signifikant geringere *Trefferrate T* (Anzahl richtig erkannter Wörter). Auch in einer weiteren Studie von Trueblood (1994) zeigte sich neben der *Trefferrate T* auch die *Lernsumme LS* (Gesamtmenge erinnerter Wörter in fünf Lerndurchgängen) als sensitiv gegenüber negativen Antwortverzerrungen. Für die beiden Indikatoren *Lernsumme LS* (Grenzwert < 48 Rohwertpunkte) und *Trefferrate T* (Grenzwert < 13 Rohwertpunkte) wurden eine Sensitivität von 70% und eine Spezifität von 91.7% ermittelt.

In einem weiteren *known-group Design* mit Schädelhirntrauma-Patienten ergänzten Millis, Putnam, Adams und Ricker (1995) die beiden Parameter *Lernsumme LS* und *Trefferrate T* um die *langverzögerten Gedächtnisleistungen mit Abrufhilfe WAI II* und *Diskriminabilität d*. Die *Diskriminabilität d* gilt, unter Einbezug der Wiedererkennungstreffer und der Falsch-Positiv-Antworten, als aussagekräftigstes Maß zur Beurteilung der Rekognitionsleistung (Niemann et al., 2008). In dieser Studie wies die *Diskriminabilität d* mit 93% die höchste Klassifikationsgenauigkeit auf. Die Klassifikationsgenauigkeit der *Trefferrate T* lag bei 89%, der *langverzögerten-gestützten Abrufleistungen WAI II* bei 87% und der *Lernsumme LS* bei 83%. Die Ergebnisse von Millis et al. (1995) wurden von Sweet et al. (2000) in einer Analogstudie mit Schädelhirntrauma-Patienten und Simulationsverdacht (auffällige Werte in mindestens einem PVT, diagnostische Diskrepanzen, objektiver Anreiz) überprüft. In einer linearen Diskriminanzanalyse mit den vier *CVLT*-Parametern ergaben sich eine Sensitivität von 73.9% und eine Spezifität von 82.5%. Mit einer Klassifikationsgenauigkeit von 78.9% lag diese unterhalb der von Millis et al. (1995) mit 91%.

In einer Analogstudie ($N = 45$) von Demakis (2004) erzielten die experimentellen Simulanten ES signifikant schlechtere Gedächtnisleistungen als die Kontrollgruppe KG in den vier Parametern *Lernsumme LS* ($d = 1.2$), *Trefferrate T* ($d = -1.3$), *langverzögerte-gestützte Abrufleistungen WAI II* ($d = -1.4$) und *Diskriminabilität d* ($d = -1.4$). Der Autor bezog die Schwellenwerte von Millis et al. (1995) und Trueblood (1994) in seine Analysen ein. Die Klassifikationsraten variierten zwischen 72% mit den Schwellenwerten von Millis et al. (1995) und 83% mit den Werten von Trueblood (1994). Die Spezifität rangierte zwischen 74% - 100%, die Sensitivität zwischen 46% - 73%. Eine Übersicht der Schwellenwerte der *CVLT-I*-Parameter ist bei Suhr und Barrash (2007) zu finden.

Im Rahmen der Neuauflage des *CVLT* im Jahr 2000 untersuchten Greve, Curtis, Bianchini und Ord (2009) die Generalisierbarkeit der eingebetteten Parameter des *CVLT-I* für die zweite Auflage des Verfahrens, den *CVLT-II* (Delis et al., 2000). Bei unverändertem Aufbau wurden im *CVLT-II* u.a. diverse Auswertungsparameter (Rekognitionswerte, Clusterwerte), einschließlich eines Alternativwahltest zur Beurteilung der Anstrengungsbereitschaft, ergänzt. Der non-parametrische Kennwert *Diskriminabilität* wurde durch den parametrischen Signaldetektionsindex *Diskriminabilität d'* ersetzt. In der *known-group Studie* von Greve und Kollegen (2009) mit einer gematchten Patientenstichprobe (Schädelhirntrauma, Schmerz) zeigten beide Gruppen (mit bzw. ohne objektivierbaren Anreiz zur Symptomverstärkung) stärkere Beeinträchtigungen im *CVLT-II* als im *CVLT-I* und es wurden unterschiedliche Klassifikationsstatistiken ermittelt. Demzufolge können die bisher ermittelten Schwellenwerte nicht uneingeschränkt auf den *CVLT-II* übertragen werden.

Die eingebetteten Parameter des *CVLT-II* untersuchten auch Wolfe et al. (2010) mit 124 Schädelhirntrauma-Patienten mit moderater bis schwerer Symptomatik sowie 77 Schädelhirntrauma-Patienten mit leichter Symptomatik und einer wahrscheinlichen oder sicheren SND nach Slick. Die Vorhersagegenauigkeit der Gruppenzugehörigkeit anhand des *CVLT-II* ermittelten die Autoren aus einer Kombination der Methode des *Bayesian model averaging (BMA)* und einem logistischen Regressionsmodell. Mittels des Bayes'schen Durchschnitts konnten die a posteriori Wahrscheinlichkeiten der eingebetteten Parameter des *CVLT-II* berechnet werden, die von den Autoren zur Variablenselektion herangezogen wurden. Von den 18 initial ausgewählten Kennwerten wurden die Parameter mit den höchsten a posteriori Wahrscheinlichkeiten ermittelt. Diese drei Kennwerte (*Diskriminabilität d'*, *Verzögerte Freie Wiedergabe II VFW II*, *nonparametrische Diskriminabilität DW*) wurden in eine schrittweise binär-logistische Regressionsanalyse als Prädiktoren eingeschlossen. Die Wahrscheinlichkeit der Gruppenzugehörigkeit (*p* Antwortverzerrungen) ermittelten die Autoren mit folgender Regressionsgleichung:

$$p \text{ Antwortverzerrungen} = \frac{e [-1.092 + 0.32(\text{Rohwert VFW II}) - 0.99(\text{Rohwert } d') - 0.693(z - \text{Wert DW})]}{1 + e [-1.092 + 0.32(\text{Rohwert VFW II}) - 0.99(\text{Rohwert } d') - 0.693(z - \text{Wert DW})]}$$

Mit einer Sensitivität von 66% und einer Spezifität von 84% konnte die Wahrscheinlichkeit der Gruppenzugehörigkeit vorhergesagt werden. In einer ROC-Analyse ermittelten die Autoren eine Sensitivität von 49% und eine Spezifität von 90% mit einem optimalen Schwellenwert der Gruppenvorhersage von $p = 0.624$.

Donders und Strong (2011) replizierten die regressionsanalytische Vorgehensweise von Wolfe et al. (2010) mit 100 Schädelhirntrauma-Patienten. Sie gliederten die Stichprobe in Patienten mit externalem Anreiz (finanzielles Kompensationsverfahren), Patienten mit psychosozialen Schwierigkeiten und Patienten ohne zusätzlich erschwerende Faktoren. 24 Patienten zeigten ein auffälliges Ergebnis im *Word Memory Test (WMT)*, so dass eine a priori Wahrscheinlichkeit für das Vorliegen negativer Antworttendenzen von 24% angenommen wurde. Anhand der Regressionsgleichung von Wolfe et al. (2010) wurden unter Einbezug der Wahrscheinlichkeit $p = 0.50$ und $p = 0.624$ moderate Sensitivitätswerte erzielt, die Spezifität lag in beiden Fällen jedoch unter 90%. Durch das Erhöhen des Schwellenwerts auf einen konservativeren Wert von $p = 0.71$ konnte die beste Klassifikationsrate (81%) mit einer Sensitivität von 33.33% und einer Spezifität von 96.05% erreicht werden. Im Vergleich zum *WMT* war die Klassifikation anhand des *CVLT-II* durch häufigere Falsch-Positiv-Entscheidungen gekennzeichnet.

Auch Schutte, Millis, Axelrod und VanDyke (2011) analysierten mit einer neurologisch-psychiatrischen Patientenstichprobe ($N = 124$) eingebettete Gedächtnisparameter. Die Autoren kombinierten den Rohwert des Alternativwahltests und den z -Wert *Lernmenge Durchgang 5 (DG5)*: Anzahl erinnerter Wörter nach dem 5. Lerndurchgang) des *CVLT-II* mit zwei weiteren Kennwerten aus neuropsychologischen Standardverfahren. Als Kriterium für suboptimale Anstrengungsbereitschaft wurde der *Medical Symptom Validity Test (MSVT)* herangezogen, in dem 26% der Patienten auffällige Werte erzielten. Eine ROC-Analyse ergab für die Parameter-Kombination eine hervorragende Wahrscheinlichkeit korrekter Klassifizierungen ($AUC = .84$, $CI = .75 - .92$). In einem logistischen Regressionsmodell mit den vier Parametern wurden eine Sensitivität von 59%, eine Spezifität von 95%, eine positive Vorhersagevalidität von 79% und eine negative Vorhersagevalidität von 87% ermittelt. Der Kappa-Koeffizient als Maß der Interrater-Reliabilität lag bei $\kappa = .59$, $p < .000$. Die diagnostische Übereinstimmung der eingebetteten Parameter und des PVT war somit moderat bis gut.

Whiteside et al. (2015) analysierten die eingebetteten Parameter verschiedener kognitiver Funktionsbereiche (Gedächtnis, Sprache und Aufmerksamkeit) bei Patienten mit Schädelhirntraumata mit moderaten bis schweren kognitiven Beeinträchtigungen und Patienten mit leichten Schädelhirntraumata. Die Patienten mit leichten Schädelhirntraumata wurden in zwei weitere Gruppen unterteilt: Patienten mit verminderter Anstrengungsbereitschaft ($n = 66$) und Patienten mit unauffälliger Anstrengungsbereitschaft

($n = 91$). Mittels einer MANOVA wurden statistisch bedeutsame Gruppendifferenzen ($\eta^2 = 0.21$) in den *CVLT-II*-Parametern *Trefferrate T*, *Falsch-Positive FP* und *Lernmenge Durchgang 5 DG5* ermittelt. Die Patienten mit verminderter Anstrengungsbereitschaft präsentierten die schlechtesten Testleistungen, während sich die beiden anderen Gruppen nicht signifikant voneinander unterschieden. Mit einer Kombination der beiden *CVLT-II*-Parameter *Falsch-Positive FP* und *Trefferrate T* und dem *Test of Attention (BTA)*: Schretlen, Bobholz & Brandt, 1996) konnte in schrittweisen logistischen Vorhersagemodellen am zuverlässigsten zwischen den Patienten mit verminderter Anstrengungsbereitschaft und den Patienten mit moderaten bis schweren kognitiven Beeinträchtigungen unterschieden werden (Nagelkerke's $R^2 = .45$). Zur Berechnung der Trennschärfe der einzelnen Parameter wurden ROC-Analysen durchgeführt. Für die *CVLT-II*-Kennwerte wurden jeweils akzeptable AUC-Werte für die *Trefferrate T* (AUC = .79) und die *Lernmenge Durchgang 5 DG5* (AUC = .72) und eine unzureichende Differenzierung für die *Falsch-Positive FP* (AUC = .67) ermittelt. Die Kombination *Brief Test of Attention*, *Falsch-Positive FP* und *Trefferrate T* erzielte eine ausgezeichnete AUC = .85 (Sensitivität 54%, Spezifität 90%).

Aus den bisherigen Forschungsarbeiten geht somit hervor, dass sich insbesondere die quantitativen Parameter der verbalen Lern- und Merkfähigkeit wie die Lernsumme, die verzögerten Abruf- und die Rekognitionsleistungen zur Erfassung negativer Antwortverzerrungen eignen. Inwieweit diese Ergebnisse auf die deutsche Version des *CVLT* übertragbar sind, wurde bislang empirisch noch nicht überprüft.

Zusammenfassend wurden diverse Untersuchungen mit den beiden englischsprachigen Versionen des *CVLT* zur Analyse eingebetteter Parameter durchgeführt. Mit dem *CVLT-I* konnten negative Antwortverzerrungen anhand der Rohwerte der *Lernsumme LS*, *Trefferrate T*, *langverzögerten Gedächtnisleistungen mit Abrufhilfe WAI II* und *Diskriminabilität d* erfasst werden. Im *CVLT-II* wurden weitere Prädiktoren wie das Alternativantwortverfahren und die Kennwerte *Lernmenge Durchgang 5 DG5*, *Diskriminabilität d'*, *nonparametrische Diskriminabilität DW*, *verzögerte freie Wiedergabe II VFW II* und *Falsch-Positive FP* ermittelt. Bislang wurden diese Ergebnisse noch nicht mit der deutschsprachigen Version des *CVLT* repliziert. Insbesondere die Kombination der *CVLT* Parameter mit Kennwerten anderer kognitiver Funktionsbereiche erscheint als vielversprechendes diagnostisches Vorgehen.

2.9.2 Reaktionszeiten als eingebettete Parameter

Seit einigen Jahren werden auch Reaktionszeiten und deren Standardabweichungen in Aufmerksamkeits- und Konzentrationstests zur Beurteilung der Anstrengungsbereitschaft herangezogen (Bodenburg, 2014; Fiene et al., 2015; Reicker, 2008; Willison & Tombaugh, 2006). In der neuropsychologischen Praxis zählen Aufmerksamkeitsstörungen zusammen mit Gedächtnisstörungen zu den häufigsten Beeinträchtigungen in Folge von Hirnschädigungen und gehen oftmals mit psychiatrischen Störungsbildern einher. In Abhängigkeit der Lokalisation und Ätiologie können unterschiedliche attentionale Funktionen beeinträchtigt sein, die auch in der „Leitlinie Diagnostik und Therapie von Aufmerksamkeitsstörungen“ von Sturm et al. (2009) aufgeführt werden: allgemeine Wachheit (Alertness), dauerhafte Aufmerksamkeitszuwendung (Daueraufmerksamkeit, Vigilanz), räumliche Ausrichtung des Aufmerksamkeitsfokus, Aufmerksamkeitsflexibilität, selektive (fokussierte) oder geteilte Aufmerksamkeit, Wechsel des Aufmerksamkeitsfokus. Beeinträchtigungen sind bei fast allen Störungsbildern zu beobachten, an denen das zentrale Nervensystem beteiligt ist. Je nach Läsion können die attentionalen Beeinträchtigungen spezifischer oder globaler Natur sein. Nach Sturm et al. (2009) sind häufige Ursachen zerebrovaskuläre Erkrankungen, Schädelhirntraumata, multiple Sklerose, neurodegenerative Erkrankungen und psychiatrische Störungen (v.a. Schizophrenie, affektive Störungen und ADHS). Bei der Diagnostik von Aufmerksamkeitsstörungen gelten insbesondere die Reaktionszeiten und Häufigkeit von Fehlreaktionen als aussagekräftig. Beeinträchtigungen der basalen Aufmerksamkeit zeigen sich primär in verlangsamten Reaktionszeiten, die auch Defizite in komplexen Aufmerksamkeitsfunktionen mitbedingen können.

Für die Analyse von Reaktionszeiten im Rahmen der BV spricht nach Reicker (2008), dass sich der Prozess der Informationsverarbeitung durch das absichtliche Generieren inkorrektur Antworten wahrscheinlich verlangsamt und sich so in längeren Bearbeitungszeiten widerspiegeln sollte. Darüber hinaus unterschätzen Personen mit Simulationsintention die Reaktionszeiten von Patienten mit tatsächlichen kognitiven Beeinträchtigungen (Resnick, 1988), was zur ausgeprägten Verlangsamung beitragen kann. Rees et al. (1998) betonen, wie Cercy, Schretlen und Brandt (1997), die Dauer des Entscheidungsprozess, der mit negativen Antwortverzerrungen einhergeht. Sie führen die Verlangsamung darauf zurück, dass Personen mit Simulationsintention zunächst die Aufgaben korrekt für sich lösen und danach entscheiden müssen, wie sie die falschen Antworten präsentieren. Diese Annahme fanden Bolan et al. (2002) im postexperimentellen Fragebogen ihrer Untersuchung bestätigt. Die experimentellen

Simulanten (ES) beschrieben, dass die Unschlüssigkeit darüber, wie sie Defizite vortäuschen sollten, zur Verlangsamung führte. Neben den Reaktionszeiten selbst betont Bodenburg (2014) auch die diagnostische Relevanz der Standardabweichungen der Reaktionszeiten. Da sich die Konsistenz des Antwortverhaltens in Reaktionszeitaufgaben durch die Standardabweichungen abbilden lässt, können Standardabweichungen als Indikatoren zur Beurteilung der Anstrengungsbereitschaft herangezogen werden. Computergestützte Aufmerksamkeitstests ermöglichen die Messung von Reaktionszeiten in Millisekunden, so dass auch kleinste Leistungsdifferenzen aufgezeichnet werden können.

2.9.2.1 Reaktionszeiten in PVT

Zunächst wurden Reaktionszeiten nur in PVT berücksichtigt. Rose, Hall und Szalda-petree (1995) und Rees et al. (1998) waren die ersten, die in Analogstudien neben der Anzahl richtiger Antworten auch die Reaktionszeiten analysierten. Sie bestätigten die Annahme, dass negative Antwortverzerrungen mit verlangsamten Reaktionszeiten einhergehen können. Auch in der Analogstudie von Rees et al. (1998) benötigten die experimentellen Simulanten ES im computergestützten *Test of Memory Malingering (TOMM)* länger für korrekte Antworten als die Kontrollgruppe KG. Bolan et al. (2002) konnten dieses Ergebnis replizieren und mit dem *Amsterdam Short Term Memory Test – C (ASTM-C: Schagen et al., 1997)* generalisieren. Auch bei Strauss et al. (1999) und Grote et al. (2000) zeigte sich eine Verlangsamung im *Victoria Symptom Validity Test (VSVT: Slick, Hopp, Strauss & Thompson, 1997)*, wobei die Reaktionszeitverzögerungen deutlicher bei schwierigeren Testitems hervortraten. In einer darauffolgenden Untersuchung von Strauss et al. (2002) präsentierten die experimentellen Simulanten ES neben langsameren Reaktionszeiten auch eine größere Variabilität der Reaktionszeiten im Vergleich zur gesunden Kontrollgruppe KG.

2.9.2.2 Reaktionszeiten in Standardverfahren

Die Analogstudie von Steck et al. (2000) untersuchte mit jeweils 30 neurologischen Patienten und gesunden Probanden den Zusammenhang zwischen Reaktionszeiten und Aufgabenkomplexität. Die gesunden Probanden erhielten drei verschiedene Untersuchungsinstruktionen. Am ersten Erhebungszeitpunkt bearbeiteten die gesunden Probanden die Aufgaben jeweils mit uneingeschränkter Anstrengungsbereitschaft als Kontrollgruppe KG. Am zweiten Erhebungszeitpunkt wurden sie als naive Simulanten, ohne detaillierte Vorinformationen instruiert, am letzten Erhebungszeitpunkt als aufgeklärte Simulanten zur minimalen Reaktionsverzögerung. Zu allen Messzeitpunkten bearbeiteten die Probanden

einfache sensomotorische Reaktionszeitmessungen und einen komplexeren Wahlversuch am *Wiener Reaktionsgerät* (WRG; Schuhfried, 1991). Im Vergleich zur Patientengruppe erzielten die naiven Simulanten langsamere Reaktionszeiten, wobei sich bei dem komplexeren Wahlversuch eine geringere Verlangsamung zeigte. Die hochsignifikanten Effekte zeigten sich auch bei den aufgeklärten ES in allen drei Subtests.

Auch Willison und Tombaugh (2006) untersuchten die Reaktionszeiten in einer Analogstudie mit experimentellen Simulanten ES, einer Kontrollgruppe KG und Schädelhirntraumapatienten. Die experimentellen Simulanten ES zeigten neben langsameren Reaktionszeiten auch eine höhere Reaktionsvariabilität und eine größere Anzahl an Fehlreaktionen im *Computerized Tests of Information Processing* (CTIP; Tombaugh & Rees, 2000). In einer weiteren Analogstudie ($N = 40$) mit dem *Computerized Tests of Information Processing* (CTIP) untersuchte Reicker (2008) den Zusammenhang zwischen Aufgabenkomplexität und Reaktionszeiten sowie deren Variationskoeffizienten (Standardabweichung/Mittelwert $\times 100$). Die Zunahme der Aufgabenkomplexität ging dabei mit einer hypotesenkonformen Verlangsamung einher. Des Weiteren erzielten die experimentellen Simulanten ES im Vergleich zur Kontrollgruppe KG langsamere Reaktionszeiten ($\eta^2 = .63$) und höhere Variationskoeffizienten ($\eta^2 = .58$). Die Klassifikationsgüte der Reaktionszeitparameter wurde anhand der Gruppenunterschiede der Mittelwerte bestimmt und mit den Trennwerten des *Test of Memory Malingering* (TOMM) verglichen. Die Spezifität des TOMM lag bei 100%, die Sensitivität bei 77%. Die Vorhersage der Gruppenzugehörigkeit anhand der Reaktionszeiten der einfachen Reaktionszeitaufgabe war mit einer Spezifität von 100% und einer Sensitivität von 87% dem TOMM überlegen. Für den Variationskoeffizient ergaben sich eine Spezifität von 97% und eine Sensitivität von 67%.

In einem *known-group Design* von Bodenburg (2014) wurde die intraindividuelle Reaktionszeitvariabilität des Subtests *Alertness* der *Testbatterie zur Aufmerksamkeitsprüfung* (TAP; Zimmermann & Fimm, 2002) untersucht. Der TAP-Subtest *Alertness* erfasst basale Aufmerksamkeitsfunktionen mit einfachen Reaktionszeitmessungen ohne Warnton (*tonische Alertness*) bzw. mit Warnton (*physische Alertness*) jeweils in blockweiser Vorgabe. Die Stichprobe mit 91 Schädelhirntrauma-Gutachtenpatienten wurde je nach Anstrengungsbereitschaft im *Word Memory Test* (WMT) in zwei Gruppen unterteilt (vermindert vs. unauffällig). In einer multivariaten Kovarianzanalyse mit den Standardabweichungen der Reaktionszeiten als abhängige Variable und diversen Prädiktoren zeigte sich ein statistisch

bedeutsamer Zusammenhang zwischen den *WMT*-Werten und der Größe der Standardabweichungen bei der *tonischen* und der *phasischen Alertness*.

Des Weiteren wurden die Kennwerte der *TAP Alertness* von Fiene et al. (2015) untersucht. Neben den Reaktionszeiten und deren Standardabweichungen bezogen die Autoren auch die Reaktionsauslassungen und Antizipationsfehler (vorzeitige Reaktion beim Warnton statt des Kreuzes) ein. In dem Untersuchungsdesign wurden, wie bei Steck et al. (2000), gesunde Probanden jeweils an drei Messzeitpunkten zur maximalen Anstrengungsbereitschaft (Standard), zur experimentellen Simulation (naive Simulanten) und zur minimalen, weniger offensichtlichen und gleichmäßigen Verlangsamung (informierte Simulanten) instruiert. Die Gruppenunterschiede der Rohwerte wurden anhand nichtparametrischer Verfahren bestimmt. Die Simulanten beider Gruppen zeigten langsamere und variabelere Reaktionszeiten als die Probanden der Standardbedingung. Die informierten Simulanten erzielten im Vergleich zu den naiven Simulanten wiederum schnellere Reaktionszeiten und geringere Standardabweichungen. Während keine Gruppenunterschiede in den Reaktionsauslassungen bestanden, wiesen die naiven Simulanten mehr Antizipationen auf. Die Analogstudie wurde durch zwei Patientengruppen (multiple Sklerose, *Mild Cognitive Impairment MCI*) ergänzt. Die naiven Simulanten unterschieden sich signifikant in den Reaktionszeiten (Mediane), Standardabweichungen und Antizipationsfehlern von beiden Patientengruppen, während die informierten Simulanten ausschließlich langsamere Reaktionszeiten als die Patienten mit multipler Sklerose erzielten. Zur Differenzierung zwischen Patienten und Simulanten wurde für jeden Leistungsparameter ein Grenzwert berechnet und zwischen 94% und 97% der Patienten korrekt klassifiziert. Die Verwendung der Grenzwerte führte zu keiner richtigen Klassifikation (Richtig-Positiv) der informierten Simulanten. Die Anzahl richtiger Klassifikationen der naiven Simulanten lag bei 53%, wobei der Trennwert von mindestens vier Antizipationsfehlern mit einer Sensitivität von 59% die beste Klassifikationsrate ermöglichte. Demzufolge eignen sich die Trennwerte der Leistungsparameter der basalen Aufmerksamkeitsmessung zur Erfassung eindeutiger Verzerrungstendenzen, für subtile Antwortverzerrungen erweist sich dieses Vorgehen bislang jedoch als nicht sensitiv genug. Eine Verbesserung der Sensitivität konnte durch eine Kombination der Prädiktoren Reaktionszeiten und Antizipationen in einem logistischen Regressionsmodell erzielt werden (Sensitivität 76.5%, Spezifität 97.2%).

Aus den Untersuchungen zu den basalen Aufmerksamkeitsfunktionen geht hervor, dass die Reaktionszeiten und Standardabweichungen im *TAP*-Untertest *Alertness* einen relevanten

diagnostischen Beitrag zur Erfassung negativer Antwortverzerrungen liefern können. Inwieweit sich die Ergebnisse auf andere *TAP*-Untertests und damit auf weitere Aufmerksamkeitsfunktionen generalisieren lassen, wurde noch nicht empirisch überprüft. Bislang wurden negative Antwortverzerrungen in komplexeren Aufmerksamkeitsfunktionen in den beiden Studien von Steck et al. (2000) und Reicker (2008) untersucht. Mit zunehmender Aufgabenkomplexität nahmen im Vergleich zu einfacheren Aufgaben die negativen Antwortverzerrungen ab.

Neben Gedächtnisparametern eignen sich auch Aufmerksamkeitsparameter zur Erfassung verminderter Anstrengungsbereitschaft. Während die Reaktionszeiten zunächst nur in PVT berücksichtigt wurden, haben sich in aktuellen Studien auch die Reaktionszeiten und deren Standardabweichungen in Aufmerksamkeits- und Konzentrationstests als vielversprechende Strategie zur BV herausgestellt. Insbesondere die basalen Aufmerksamkeitsfunktionen des *TAP*-Untertests *Alertness* können als eingebettete Parameter mit Schwellenwerten oder logistischen Vorhersagemodellen die Diagnostik verminderter Anstrengungsbereitschaft ergänzen.

2.10 Diagnostische Güte und Anzahl verwendeter Methoden

Mit der kombinierten Anwendung unterschiedlicher Verfahren geht auch die Frage einher, inwieweit durch mehrere Indikatoren die diagnostische Güte zur Erfassung negativer Antwortverzerrungen beeinflusst werden kann. Um den Einfluss der Anzahl verwendeter Parameter zu erfassen, schlägt Larrabee (2008) die Verkettung von Wahrscheinlichkeitsverhältnissen (*Likelihood Ratios, LR*) zur Ermittlung der a posteriori Wahrscheinlichkeiten negativer Antwortverzerrungen vor. Das positive Wahrscheinlichkeitsverhältnis $LR+$ entspricht dabei dem Verhältnis der Sensitivität eines Kennwerts zu dessen Falsch-Positiv-Rate (1-Spezifität). In einer Studie von Chafetz (2011) lag die Prävalenzrate für negative Antwortverzerrungen in einer Stichprobe von 161 Gutachtenpatienten bei 38.5%. In Abhängigkeit der Klassifikationsgüte des Verfahrens zur BV stieg die a posteriori Wahrscheinlichkeit bei einem auffälligen Kennwert auf 84% bis 91%. Bei zwei auffälligen Kennwerten lag die a posteriori Wahrscheinlichkeit zwischen 93% und 97%. Mit drei auffälligen BV-Kennwerten wurde eine a posteriori Wahrscheinlichkeit von 99% erzielt. Demzufolge kann durch die Kombination mehrerer Indikatoren die Sensitivität erhöht werden, ohne dabei die Spezifität zu vermindern, vorausgesetzt die Messwerte sind voneinander

unabhängig und weisen geringe Interkorrelationen auf (Bashem et al., 2014; Berthelson, Mulchan, Odland, Miller & Mittenberg, 2013; Larrabee, 2008). Berthelson et al. (2013) untersuchten in einer Metaanalyse die Interkorrelationen verschiedener BV-Parameter, die im Durchschnitt bei $r = .31$ lagen. Mit einer Monte Carlo Simulation ermittelten die Autoren die Rate Falsch-Positiver-Entscheidungen aus der Anzahl auffälliger BV-Parameter im Verhältnis zur Anzahl verwendeter Parameter. Dabei berücksichtigten die Autoren neben den Interkorrelationen auch eine Fehlerrate für Falsch-Positiv-Entscheidungen der Testverfahren. Aus ihren Ergebnissen schlossen die Autoren, dass mit einer Zunahme der Anzahl verwendeter Testverfahren bei konstanter Anzahl an auffälligen Parametern die Rate an Falsch-Positiv-Entscheidungen zunimmt. Während die Rate an Falsch-Positiv-Entscheidungen bei der Verwendung von zwei Testverfahren bei 2.2% lag, stieg sie mit fünf Testverfahren auf 11.5% und mit sieben Testverfahren auf 17.5% (bei 10% Fehlerrate der Testverfahren). Die vorhergesagten Raten der Falsch-Positiv-Entscheidungen von Berthelson et al. (2013) wurden von mehreren Autoren wie Larrabee (2014), Bashem et al. (2014) sowie Davis und Millis (2014) mit den klinischen Daten ihrer Untersuchungen verglichen. Die Vorhersagewerte konnten die Autoren nicht bestätigen, vielmehr gingen sie von einer Überschätzung Falsch-Positiv-Entscheidungen aus. Durch die Verwendung von mehr als zwei Verfahren konnten Bashem et al. (2014) jedoch nur eine geringe bis moderate Verbesserung der diagnostischen Güte erzielen. Larrabee (2014) befürwortet als Implikation für die Praxis die Verwendung von mindestens zwei Verfahren.

2.11 Diagnostische Grenzen der Beschwerdvalidierung

In diesem Kapitel werden die diagnostischen Grenzen der BV beschrieben. Zunächst werden die Konsequenzen gezielter Vorabinformationen und Coaching dargestellt. Ferner werden die Auswirkungen tatsächlicher kognitiver Störungen, insbesondere demenzieller Störungsbilder, vertieft. Abschließend wird die Rolle der BV-Verfahren im diagnostischen Prozess der Simulation kognitiver Störungen erläutert.

2.11.1 Beschwerdvalidierung und Coaching

Unter Coaching verstehen Suhr und Gunstad (2007) den gezielten Einsatz von Informationen, wodurch ein glaubhafter Eindruck kognitiver Beeinträchtigungen entstehen soll. Die Autoren differenzieren zwischen der Ankündigung des Einsatzes von BV-Methoden (*Vorwarnen*), Informationen über typische Symptome des vorzutäuschenden Störungsbildes

(*Symptominformation*) und Informationen über die neuropsychologischen Testverfahren (*Testinformation*). Das entsprechende Wissen kann aus unterschiedlichen Informationsquellen stammen. In Nordamerika stellt das sogenannte „Anwalt-Coaching“ (Merten, 2005, S. 39, Spalte 2, Z. 20) im Vorfeld einer neuropsychologischen Begutachtung keine Seltenheit dar. Merten (2005) schließt eine entsprechende Entwicklung im deutschen Sprachraum nicht aus. Als weitere Informationsquellen dienen laut Suhr und Gunstad (2007) neuropsychologische Untersuchungen selbst, da Patienten durch Testerfahrung und Ergebnisrückmeldungen ihre Testperformanz in einer erneuten Untersuchung entsprechend beeinflussen können. Der Austausch unter den Patienten und der einfache Informationszugang des Internets können ebenfalls als Coaching fungieren. Im Rahmen einer Internetrecherche mit der Suchmaschine *Google* schätzten Bauer und McCaffrey (2006) 26% der ermittelten Internetseiten als moderat bis stark gefährdend für die Testsicherheit von gängigen PVT ein (Suhr & Gunstad, 2007). Ziel der Forschung ist die Entwicklung robuster Strategien gegenüber Coaching (Rogers, 1997; Rüsseler et al., 2008). In diversen Studien konnten mit Symptominformationen keine glaubhafteren kognitiven Defizite vorgetäuscht werden (Gorny & Merten, 2007; Rüsseler et al., 2008; Weinborn et al., 2012). Auch das Vorwissen durch berufliche Expertise führte in einer Studie von Strauss et al. (2002) nicht zu besseren Leistungen in PVT im Vergleich zu naiven Simulanten ohne Vorwissen. Die glaubhaftesten Defizite wurden mit der Kombination aus Symptom- und Testinformation vorgetäuscht, wobei Testinformationen zu glaubhafteren Defiziten führten als Symptominformationen (Weinborn et al., 2012). Der Einfluss von Vorwarnungen auf die Testperformanz wurde bislang nicht eindeutig geklärt. Während Vorwarnungen in einer Untersuchung von Greub und Suhr (2006) keinen Effekt auf die Testperformanz hatten, zeigten vorgewarnte Probanden in der Studie von Gorny und Merten (2007) überzeugendere Beeinträchtigungen als naive Simulanten. In der deutschsprachigen Begutachtungspraxis warnen 34% der Gutachter nahezu jeden Patienten über die Erfassung der Anstrengungsbereitschaft vor. Im klinischen Kontext liegt die Rate regelmäßiger Vorwarnungen mit 13% deutlich niedriger. Eine Aufforderung zur maximalen Anstrengungsbereitschaft geben 46% der Diagnostiker im klinischen Setting und 71% im Gutachtenkontext (Dandachi-FitzGerald et al., 2013). Sowohl im Positionspapier der NAN (Bush et al., 2005) als auch in der Konsenserklärung der AACN (Heilbronner et al., 2009) werden Neuropsychologen dazu angehalten, Patienten zur optimalen Anstrengungsbereitschaft zu ermutigen. Allerdings liegen diesbezüglich keine Richtlinien vor, insbesondere im Umgang mit Patienten, die negativen Antwortverzerrungen zeigen (Martin et al., 2015).

2.11.2 Beschwerdvalidierung und kognitive Dysfunktionen

Eine weitere grundlegende Einschränkung aller standardisierten Methoden der BV sind schwere kognitive Beeinträchtigungen (Rienstra, Klein Twennaar & Schmand, 2013; Schmand et al., 2005b). Obwohl PVT auf der Annahme basieren *nicht sensitiv* gegenüber kognitiven Störungen zu sein (Tombaugh, 1997; Willis et al., 2011), können kognitive Dysfunktionen zu auffälligen BV-Parametern und somit zu Fehlinterpretation führen. Eine Konfundierung von verminderter Anstrengungsbereitschaft und tatsächlichen kognitiven Beeinträchtigungen geht mit einer Abnahme der Spezifität und mit einem steigenden Risiko von Falsch-Positiv-Klassifikationen einher (Merten, Bossink, et al., 2007). In der neuropsychologischen Praxis im deutschen Sprachraum wird das Vorliegen einer offenkundigen kognitiven Störung als wichtigstes Argument gegen die Verwendung von PVT angesehen (Dandachi-FitzGerald et al., 2015). Die Grenzen der meisten PVT sind jedoch nicht eindeutig definiert und rangieren auf einem Kontinuum von leichtgradigen bis schwerwiegenden kognitiven Beeinträchtigungen (Merten, Bossink, et al., 2007). Um die Klassifikationsgenauigkeit der PVT bei kognitiv beeinträchtigten Patienten zu spezifizieren, wurden Demenz-Patienten in diversen Studien untersucht, beispielsweise bei Green, Montijo und Brockhaus (2011); (Merten, Bossink, et al., 2007); Rienstra et al. (2013); Teichner und Wagner (2004); Walter, Morris, Swier-Vosnos und Pliskin (2014).

2.11.2.1 Störungsbild Demenz

In den beiden Klassifikationssystemen ICD-10 und DSM-5 wird Demenz als eine erworbene Beeinträchtigung des Gedächtnisses und Einschränkungen des kognitiven Leistungsniveaus definiert. Zudem ist mindestens eine weitere kognitive Funktion (v.a. Sprache, Orientierung, Konzentration, abstraktes Denken und Urteilsvermögen) betroffen. Neurodegenerative Entwicklungen können auch mit Veränderungen der Persönlichkeit, der emotionalen Kontrolle und Motivation, der Psychomotorik und des Sozialverhaltens einhergehen, wohingegen eine Bewusstseinsstrübung als Ausschlusskriterium gilt. Die Beeinträchtigungen führen zu verminderter Alltagskompetenz, die ein notwendiges Kriterium zur Diagnosestellung ist (Jahn & Werheid, 2015; Möller, 2013). Epidemiologisch steigt die Auftretenswahrscheinlichkeit einer Demenz mit zunehmendem Alter. Die Prävalenz demenzieller Syndrome liegt laut Bickel (2012) bei ca. 2% in der Altersspanne von 65 bis 70 Jahren und steigt auf über 20% ab einem Lebensalter von 85 Jahren. Unter Berücksichtigung leichter Demenzformen ist insgesamt mit einer Prävalenzrate von 10% ab dem 65. Lebensjahr auszugehen (Möller, 2013). Neben unterschiedlichen Formen geht man auch von vielfältigen

Ätiologien aus. Die beiden wichtigsten Ursachen unterteilen Jahn und Werheid (2015) in neurodegenerative Veränderungen (z.B. Morbus Alzheimer, Morbus Parkinson) und vaskuläre Ereignisse (z.B. Multi-Infarkt-Demenz). Auch ernährungsbezogene Mangelerscheinungen wie ein Vitamin B12-Mangel, internistische Krankheitsbilder (z.B. Hypertonie) und langjähriger Substanzabusus (z.B. Alkoholismus) stellen mögliche Ursachen dar. Neben den vaskulären Demenzen, die durch Veränderungen der Hirngefäße entstehen, sind als weitere Formen die gemischte Demenz, die frontotemporale Demenz und die Lewy-Körperchen-Demenz zu ergänzen. Die häufigste Form ist die Demenz vom Alzheimer-Typ (60%), deren neuronale Verfallserscheinungen (Neurodegeneration) im Temporal- und Parietallappen deutlich werden. Dieser Atrophie liegen neben Proteinablagerungen, im Sinne von β -Amyloid zwischen den Neuronen (Plaques), auch Neurofibrillenveränderungen durch Phosphorylierung von Tau-Proteinen und eine Vermehrung der Gliazellen zugrunde (Jahn & Werheid, 2015). Die Entwicklung demenzieller Syndrome wird häufig von einer depressiven Symptomatik begleitet. Eine differentialdiagnostische Abgrenzung der beiden Störungsbilder ist, insbesondere zu Beginn, auf Grund überlappender Kriterien oftmals schwierig. Für die weiterführende Behandlung ist sie jedoch ausschlaggebend (Möller, 2013). Giebel et al. (2016) schließen aus bisherigen Studien auf eine Komorbidität von 6% und 61%. In einer Langzeitstudie von Snowden et al. (2015) zeigten Demenz-Patienten mindestens doppelt so häufig depressive Symptome wie kognitiv unbeeinträchtigte Probanden (*Odds Ratio* = 2.64). Zudem lag die Wahrscheinlichkeit bei den Demenzpatienten innerhalb von zwei Jahren eine Depression zu entwickeln bei 24.7%. Prado-Jean et al. (2010) fanden in einer Studie mit 319 Demenzpatienten eine Depressionsrate von 40% und einen moderierenden Effekt des Schweregrades der Demenz auf die affektive Symptomatik. Depressive Störungen waren weniger häufig bei Patienten mit leichteren demenziellen Symptomen.

2.11.2.2 PVT bei Patienten mit schweren kognitiven Beeinträchtigungen

Bei Vorliegen tatsächlicher kognitiver Störungen ist die Frage, ab welchem Punkt unabhängige PVT nicht mehr die Anstrengungsbereitschaft, sondern vielmehr die kognitiven Fähigkeiten erfassen, bisher noch nicht abschließend geklärt (Brockhaus & Merten, 2004). In einer Studie mit 32 geistig behinderten Personen von Brockhaus und Merten (2004) bestanden 31 Patienten den *Word Memory Test (WMT)*. Dagegen erzielten fast alle bona-fide Patienten mit Alzheimerdemenz (90% bis 95%) in einer Studie von Merten, Bossink, et al. (2007) auffällige *WMT*-Kennwerte. Zur Berücksichtigung des kognitiven Status entwickelte Green ein Demenz-Profil (*Genuine Memory Impairment Profile*). Die *WMT*-Subtests mit niedrigem

Schwierigkeitsgrad zur Erfassung der Anstrengungsbereitschaft werden durch kognitiv anspruchsvollere Subtests ergänzt. Bei Demenzpatienten ist tendenziell ein Leistungsabfall in den anspruchsvolleren Aufgaben im Vergleich zu den einfacheren Aufgaben zu beobachten (Green et al., 2011; Rienstra et al., 2013). In einer Studie von Green et al. (2011) hatten alle Demenzpatienten, die die einfachen Subtests nicht bestehen konnten eine Mittelwertdifferenz zwischen den einfachen und schwierigen Subtests von mindestens 30 Punkten. In einer Verlaufsstudie von Rienstra et al. (2013) zeigten kognitiv beeinträchtigte Patienten mit auffälligen Demenz-Profilen im *WMT* nach zwei Jahren eine stärkere Progredienz bestehender kognitiver Symptome als Patienten mit unauffälligen Demenz-Profilen. Die Sensitivität des Demenz-Profiles war mit 60% moderat, die Spezifität mit 81% bzw. 93% im Follow-up hoch. Ein alternatives Vorgehen schlagen Dean, Victor, Boone, Philpott und Hess (2009) vor. Die Autoren verglichen die Testperformanz von 214 Demenzpatienten in 18 PVT und eingebetteten Parametern mit dem Rohwert der *Mini-Mental State Examination (MMSE)*: Folstein, Folstein, McHugh & Fanjiang, 2000), einem Screeningverfahren zur Beurteilung des kognitiven Status mit einer Maximalpunktzahl von 30 Rohwertpunkten. Die Verschlechterung des kognitiven Status ging mit einem niedrigeren *MMSE*-Rohwert und einer steigenden Anzahl auffälliger BV-Indikatoren einher. Während die Demenzpatienten mit einem *MMSE* von mindesten 20 Rohwertpunkten durchschnittlich in 36% der BV-Verfahren auffällige Ergebnisse erzielten, fielen Patienten mit einem *MMSE* kleiner als 15 durch 83% der Verfahren. Auf Grund ihrer Ergebnisse setzen die Autoren das Vortäuschen kognitiver Defizite mit einem Nichtbestehen von 85% der angewandten BV-Verfahren ohne offensichtliche Beeinträchtigung der kognitiven Fähigkeiten gleich.

Merten, Bossink, et al. (2007) untersuchten in zwei unabhängigen Studien die diagnostischen Grenzen des *Amsterdam Short-Term Memory Test (ASTM)*, *Word Memory Test (WMT)* und *Test of Memory Malingering (TOMM)* mit Patienten mit kognitiven Störungen. In Studie 1 wurden 48 neurologische bona-fide Patienten in zwei Gruppen (Gruppe 1: kognitiv beeinträchtigt vs. Gruppe 2: kognitiv unauffällig) unterteilt. Die Einteilung erfolgte u.a. anhand der Kriterien von Schmand et al. (2005b) für offensichtliche kognitive Beeinträchtigungen im Kontaktverhalten wie beispielsweise dem Wiederholen bereits erzählter Informationen, fehlendes Wissen über aktuelle persönliche Umstände oder Wortfindungsschwierigkeiten. Der mittlere *MMSE* der Gruppe 1 lag bei 22.5 Rohwertpunkten ($SD = 3.1$; Spannweite = 17 - 28). In Gruppe 2 betrug der mittlere *MMSE* 28.1 Rohwertpunkte ($SD = 1.3$; Spannweite = 26 - 30). Aus den non-parametrischen Ergebnisanalysen (*U*-Test)

ging hervor, dass die beeinträchtigten Patienten der Gruppe 1 in sämtlichen PVT-Parametern eine signifikant schlechtere Testperformanz als Gruppe 2 erzielten. Aus Gruppe 2 bestanden 100% der Patienten den *TOMM* und jeweils 92% den *ASTM* und *WMT*. Dagegen bestanden lediglich die Hälfte der kognitiv beeinträchtigten Patienten aus Gruppe 1 den *WMT* (50% bis 58%) und den *ASTM* (54%). Die Klassifikationsgenauigkeit des *ASTM* lag bei 73%. Die Klassifikationsgenauigkeit des *WMT* (ohne additionalen Gedächtnissubtests) variierte zwischen 71% und 79%. Die höchste Klassifikationsgenauigkeit wurde mit dem *TOMM* (90% bis 92%) ermittelt. Die Ergebnisse verdeutlichen, dass PVT trotz kognitiver Dysfunktionen bearbeitet werden können. Dabei sind Fehlklassifikationen mit dem *WMT* und *ASTM* nicht auszuschließen. Auch in einer zweiten Studie von Merten, Bossink, et al. (2007) mit 20 Alzheimerpatienten (Durchschnittsalter = 73.5 Jahre) und 14 Kontrollpersonen (Durchschnittsalter = 76.6 Jahre) zeigten die Alzheimerpatienten signifikant schlechtere Leistungen im *ASTM*, *WMT* und *TOMM* als die Kontrollpatienten. Während alle Kontrollpatienten die drei PVT bestanden, erzielten 90% der Alzheimerpatienten auffällige Werte im *ASTM* und 90% bis 95% im *WMT* (je nach Parameter). Dagegen bestanden 50% bzw. 70% der Alzheimerpatienten die Subtests des *TOMM*. Zudem ermittelten die Autoren einen statistisch bedeutsamen Zusammenhang zwischen der Testperformanz im *ASTM* und dem Arbeitsgedächtnis, ebenso zwischen dem deklarativen Gedächtnis und dem *TOMM* und *WMT*.

Auch die Bearbeitung des *Aggravations- und Simulations-Test (AST)* beansprucht nach Eberl und Wilhelm (2007) das Arbeitsgedächtnis. Die Autoren führen im Handbuch des *AST* die Interkorrelationskoeffizienten einer kognitiv beeinträchtigten Patientenstichprobe auf. Die Anzahl richtiger Antworten im *AST (AST P)* korrelierte mit einer guten Arbeitsgedächtnisleistung ($r = 0.39$), die durch die *Digit Span backward* erfasst wurde. Zudem zeigten sich signifikante Korrelationen zwischen dem Median der Reaktionszeiten im *AST (AST Md)* und der kognitiven Informationsverarbeitungsgeschwindigkeit im *TMT-A* ($r = .53$), dem Arbeitsgedächtnis in der *Digit Span backward* ($r = -.47$) und der kognitiven Flexibilität *TMT-B* ($r = .35$). Demzufolge sind bei der Bearbeitung des *AST* auch die kognitive Informationsverarbeitungsgeschwindigkeit und die kognitive Flexibilität beteiligt. Dies wurde bislang in keiner unabhängigen Studie untersucht.

Zu den Grenzen der neuropsychologischen BV zählt Coaching über Symptome und Testverfahren, wodurch die Testperformanz gezielter manipuliert und schwieriger erfasst werden kann. Ein weiteres Risiko für Falsch-Positiv-Fehlklassifikationen stellen kognitive Störungen wie demenzielle Erkrankungen dar. In diversen Studien zeigten Patienten mit kognitiven Beeinträchtigungen trotz uneingeschränkter Anstrengungsbereitschaft auffällige Werte in verschiedenen PVT. Dabei bestehen Patienten mit leichten kognitiven Beeinträchtigungen häufiger PVT als Patienten mit schweren kognitiven Störungen. Darüber hinaus wurde mit Hilfe von kognitiv beeinträchtigten Patienten ein Zusammenhang zwischen dem Arbeitsgedächtnis und der Testperformanz im *ASTM* und im *AST* ermittelt. Ferner korrelierten die kognitive Informationsverarbeitungsgeschwindigkeit und die kognitive Flexibilität mit dem *AST*.

3 Fragestellungen und Hypothesen

Die dargestellten Studienbefunde verdeutlichen die guten diagnostischen Gütekriterien des *Word Memory Tests (WMT)* und des *Amsterdamer Kurzzeitgedächtnistests (AKGT)*. Zudem wurde der *WMT* in diversen Untersuchungen als Außenkriterium zur Validierung neuer Verfahren herangezogen. Der *Aggravations- und Simulations-Test (AST)* wurde dagegen lediglich in einer unabhängigen *known-group Studie* untersucht. Im Vergleich zum *WMT* wurden mit dem *AST* weniger Richtig-Positive-Entscheidungen getroffen.

Darüber hinaus wird Einsatz von PVT zunehmend durch eingebettete Parameter ergänzt. In bisherigen Studien mit den englischsprachigen Versionen des Gedächtnistests *California Verbal Learning Test (CVLT)* wurden diverse Parameter zur Erfassung der Anstrengungsbereitschaft ermittelt und miteinander kombiniert: *Lernmenge Durchgang 5 (DG5)*, *Lernsumme (LS)*, *verzögerte freie Wiedergabe II (VFW II)*, *langverzögerten Gedächtnisleistungen mit Abrufhilfe (WAI II)*, *Trefferrate (T)*, *Falsch-Positive (FP)* und *Diskriminabilität (d')*. Diese Parameter setzen sich vorwiegend aus mehreren Durchgängen bzw. Faktoren zusammen. Demzufolge kann der diagnostische Ansatz inkonsistenter Leistungskurven (siehe Abschnitt 2.7.3) auf die Gedächtnisleistung im *CVLT* angewandt werden. Es erscheint plausibel, dass sich Inkonsistenzen der Gedächtnisleistungen vor allem in Kennwerten widerspiegeln, die sich aus mehreren Messungen zusammensetzen und untypische Abweichungen verdeutlichen.

Darüber hinaus scheint auch die Analyse der Aufmerksamkeitsfunktionen ein geeigneter Ansatz zu sein. Neben den verlangsamten Reaktionszeiten erwiesen sich auch die intraindividuelle Reaktionszeitvariabilität (*SD*) und die Anzahl an vorzeitigen Reaktionen (Antizipationen) des Untertests *Alertness* der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* als diagnostisch relevant. Zudem wird ein Zusammenhang zwischen Reaktionszeit und der Aufgabenkomplexität angenommen. Bei komplexen Aufgaben wurden bislang weniger ausgeprägte Antwortverzerrungen beobachtet. Allerdings zeigten Patienten mit mangelnder Anstrengungsbereitschaft, die im Rahmen der diagnostischen Tätigkeit der AG Klinische und Experimentelle Neuropsychologie des Klinikums rechts der Isar der TU München untersucht wurden, neben Beeinträchtigungen in der *Alertness* auch Defizite in der *geteilten Aufmerksamkeit* und *Inhibitionsfähigkeit*. Computergestützte Aufmerksamkeits-tests wie die *TAP* mit Answererfassung durch Reaktionszeit-Tasten ermöglichen auf einfache Art und Weise, Aufmerksamkeitsdefizite vorzutäuschen. Inwiefern diese Aufmerksamkeitsfunktionen

oder eine Kombination aus basalen und komplexen attentionalen Funktionen bzw. attentionalen und mnestischen Funktionen diagnostisch weiterführend sind, wurde empirisch noch nicht untersucht.

PVT basieren auf dem Konstruktionsprinzip der verdeckten Leichtigkeit und erscheinen kognitiv anspruchsvoller als sie tatsächlich sind. Insofern ist anzunehmen, dass die subjektive Schwierigkeit eines PVT die diagnostische Güte beeinflussen kann. Demnach sollten schwierigere PVT Probanden mit verminderter Anstrengungsbereitschaft zu mehr Fehlern einladen als einfachere PVT. Ferner wurden in der Literatur die Anwendungsgrenzen der BV diskutiert, die bei Vorliegen authentischer kognitiver Dysfunktionen schnell überschritten werden können. Demenzpatienten bearbeiten PVT oft fehlerhaft, was die Gefahr von Fehlklassifikationen erhöht. Zudem erlauben Studien mit kognitiv beeinträchtigten Patientenstichproben Rückschlüsse auf die kognitiven Funktionen, die für die Bearbeitung einzelner PVT benötigt werden. So beanspruchten der *AKGT* und der *AST* vor allem das Arbeitsgedächtnis. Des Weiteren korrelierte der *AST* mit der kognitiven Flexibilität und der Informationsverarbeitungsgeschwindigkeit. Inwiefern sich die beiden Verfahren in den zugrunde liegenden kognitiven Funktionen unterscheiden, wurde bislang nicht untersucht.

In der vorliegenden Arbeit wurden sieben Fragestellungen (F) mittels zehn Hypothesen (H) bearbeitet. Es wurde eine Analogstudie mit zwei ergänzenden klinischen Stichproben durchgeführt. Die Hypothesen wurden mit den jeweiligen Stichproben überprüft und die Studie in drei Teile untergliedert. Im ersten Teil der Studie wurde die Testperformanz gesunder Probanden (experimentelle Simulanten ES vs. Kontrollgruppe KG) im *AKGT*, *AST* und *WMT* und diversen neuropsychologischen Standardverfahren (u.a. *CVLT* und *TAP*) miteinander verglichen. Darüber hinaus wurde der Zusammenhang zwischen der subjektiven Aufgabenschwierigkeit und der Testperformanz ermittelt. Anschließend wurden die Klassifikationsstatistiken der PVT und der eingebetteten Parameter der Standardverfahren berechnet. Im zweiten Teil der Studie wurde die Generalisierbarkeit der Ergebnisse aus dem Analogstudiendesign anhand einer psychiatrisch-neurologischen Stichprobe mit verminderter Anstrengungsbereitschaft (PNA) untersucht. Dazu wurden die Leistungsparameter der ES mit der psychiatrisch-neurologischen Stichprobe mit verminderter Anstrengungsbereitschaft (PNA) verglichen. Neben der Testperformanz wurden auch die Klassifikationen mit den Methoden der BV aus dem ersten Teil der Studie überprüft. Der dritte Teil umfasste den Vergleich der Testperformanz im *AST* und *AKGT* zwischen den Patienten mit verminderter Anstrengungsbereitschaft (PNA) und einer weiteren psychiatrisch-neurologischen Stichprobe

mit authentischen kognitiven Beeinträchtigungen (PKB). Dabei wurden Falsch-Positiv-Klassifikationen und der Zusammenhang zwischen der Testperformanz und den zugrunde liegenden kognitiven Funktionsbereichen untersucht.

3.1 Erster Teil der Studie: Testperformanz im experimentellen Analogdesign

F1: Bestehen Unterschiede in der Testperformanz zwischen Personen mit maximaler und verminderter Anstrengungsbereitschaft?

H1: Personen mit verminderter Anstrengungsbereitschaft erzielen in den PVT-Parametern des *AKGT*, *WMT* und *AST* signifikant schlechtere Leistungen als Personen mit maximaler Anstrengungsbereitschaft.

H2: Personen mit verminderter Anstrengungsbereitschaft erzielen in neuropsychologischen Standardverfahren signifikant schlechtere Leistungen als Personen mit maximaler Anstrengungsbereitschaft.

F2: Beeinflusst die subjektive Schwierigkeit eines PVT die Testperformanz?

H3: Personen mit verminderter Anstrengungsbereitschaft zeigen in subjektiv einfach erlebten PVT weniger Fehler als in schwierigeren PVT. Es besteht ein positiver Zusammenhang zwischen der subjektiven Schwierigkeit und der Anzahl an Fehlern in PVT.

F3: Lassen sich die Klassifikationsstatistiken der Autoren des *AST* replizieren?

H4: Sämtliche Probanden der experimentellen Analogstudie können anhand des *AST* entsprechend ihrer Gruppenzugehörigkeit korrekt klassifiziert werden.

F4: Inwiefern eignen sich verbal-mnestische und attentionale Testvariablen neuropsychologischer Standardverfahren zur Vorhersage der Gruppenzugehörigkeit?

H5: Die experimentelle Gruppenzugehörigkeit kann zuverlässig aus den *CVLT*-Parametern vorhergesagt werden.

H6: Die experimentelle Gruppenzugehörigkeit kann zuverlässig aus dem *TAP*-Untertest *Alertness* vorhergesagt werden. Auch die Parameter der *TAP*-Untertests der *geteilten Aufmerksamkeit* und *Inhibitionsfähigkeit* können zur Gruppenvorhersage herangezogen werden.

3.2 Zweiter Teil der Studie: Testperformanz der Patienten mit negativen Antwortverzerrungen (PNA)

- F5:** Bestehen Unterschiede in der Testperformanz zwischen experimentellen Simulanten (ES) und psychiatrisch-neurologischen Patienten mit negativen Antwortverzerrungen (PNA)?
- H7:** Es bestehen statistisch bedeutsame Unterschiede im *AST* zwischen den experimentellen Simulanten und der klinischen PNA-Stichprobe. Mit dem *AST* können weniger Richtig-Positiv-Entscheidungen der PNA getroffen werden.
- H8:** Es bestehen statistisch bedeutsame Unterschiede im *CVLT* und der *TAP* zwischen den experimentellen Simulanten (ES) und der klinischen PNA-Stichprobe. Mit den eingebetteten Parametern können weniger Richtig-Positiv-Entscheidungen der PNA getroffen werden.

3.3 Dritter Teil der Studie: Testperformanz der Patienten mit kognitiven Beeinträchtigungen (PKB)

- F6:** Bestehen Unterschiede im *AKGT* und *AST* zwischen den Patienten mit verminderter Anstrengungsbereitschaft (PNA) und den Patienten mit authentischen kognitiven Beeinträchtigungen (PKB)?
- H9:** Es zeigen sich keine statistisch bedeutsamen Unterschiede zwischen den beiden Patientenstichproben im *AKGT* und *AST*. Bei den Patienten mit authentischen kognitiven Beeinträchtigungen (PKB) ist von einer erhöhten Anzahl an Falsch-Positiv-Fehlklassifikationen auszugehen.
- F7:** Welche kognitiven Funktionen werden während der Bearbeitung des *AKGT* und *AST* beansprucht?
- H10:** Die Bearbeitung des *AKGT* und *AST* involviert diverse kognitive Funktionen. Es bestehen signifikante positive Korrelationen zwischen mehreren Funktionsbereichen und dem *AKGT* und *AST*.

4 Methodik: erster Teil der Studie

Die vorliegende Untersuchung wurde als Analogstudie mit gesunden Probanden konzipiert (erster Teil der Studie) und um zwei neurologisch-psychiatrische Patientenstichproben ergänzt. Ein partieller Stichprobenvergleich der experimentellen Simulanten (ES) und Patienten mit negativen Antwortverzerrungen (PNA) stellt den zweiten Teil der Studie dar. Im dritten Teil der Studie werden die Patienten mit negativen Antwortverzerrungen (PNA) einer weiteren klinischen Stichprobe, bestehend aus bona-fide Patienten mit kognitiven Beeinträchtigungen (PKB), gegenübergestellt. Ziel ist es, nach einer primären Validierung neuer BV-Methoden durch die Analogstudie, die experimentell ermittelten Ergebnisse im klinischen Kontext zu überprüfen.

In diesem Kapitel wird die experimentelle Analogstudie beschrieben. Dies umfasst die experimentelle Operationalisierung, die verwendeten psychometrischen Testverfahren, die statistischen Auswertungsschritte und die Stichprobenbeschreibung.

4.1 Studiendesign und Ablauf

In dem unifaktoriellen Versuchsplan der Analogstudie stellte die Instruktion zur Aufgabenbearbeitung die unabhängige Variable dar. Anhand der beiden Faktorstufen *maximale Anstrengungsbereitschaft* und *Simulation kognitiver Beeinträchtigung* wurde die Anstrengungsbereitschaft variiert, woraus die beiden Untersuchungsbedingungen Kontrollgruppe (KG) und experimentelle Simulanten (ES) resultierten. Durch eine randomisierte Gruppeneinteilung wurde die wichtigste Voraussetzung eines experimentellen Designs erfüllt (Hager, 1987). Die Zuteilung der Probanden erfolgte anhand von Randomisierungslisten, die mit einem Generator für Zufallszahlen erstellt wurden (Kummer, 2014). Die abhängigen Variablen waren die Testergebnisse in drei PVT und sieben neuropsychologischen Standardverfahren. Da alle Untersuchungen von der Doktorandin selbst durchgeführt wurden, konnte die Studie nicht als Blindversuch konzipiert werden.

Die Studienteilnahme beinhaltete eine etwa zweistündige neuropsychologische Untersuchung, die jeweils zusammenhängend an einem Tag durchgeführt wurde. Neben primär neuropsychologischen Leistungsvariablen und der Anstrengungsbereitschaft wurden auch soziobiographische und klinische Angaben sowie eine postexperimentelle Einschätzung der Aufgabenschwierigkeit und der eigenen Anstrengungsbereitschaft erfasst. Zu Beginn wurden alle Probanden sowohl mündlich als auch schriftlich über Sinn und Zweck der Untersuchung,

die Freiwilligkeit und die Möglichkeit eines Teilnahmeabbruchs ohne Nennung von Gründen sowie die Anonymität der Datenauswertung aufgeklärt (Anhang B). Nach einer schriftlichen Einverständniserklärung (Anhang C) wurde den Probanden eine schriftliche Rückmeldung ihrer Ergebnisse angeboten (Anhang D). Neun Probanden füllten das dafür entworfenen Kontaktblatt aus und erhielten eine entsprechende Ergebnisrückmeldung. Danach beantworteten die Teilnehmer einen Anamnesebogen (Anhang E) und lasen, je nach Untersuchungsbedingung, die entsprechende schriftliche Instruktion (Anhang F; Anhang G). Anschließend wurden die Probanden zur Wiedergabe der Instruktion mit eigenen Worten gebeten (Anhang H). Daraufhin erfolgte die Ankündigung, dass unabhängig von der Testperformanz zu zwei festgegebenen Zeitpunkten im Verlauf der Untersuchung eine Instruktionserinnerung erfolgen würde. Dann bearbeiteten die Probanden die psychometrischen Leistungstests und PVT. Die PVT folgten nicht direkt aufeinander, sondern waren als erstes und letztes Verfahren sowie in der Mitte der Untersuchung platziert. Durch die Einbettung der PVT in eine neuropsychologische Standarduntersuchung wurde eine möglichst realistische Nachbildung einer tatsächlichen Untersuchungssituation angestrebt (Merten et al., 2004). Mögliche Reihenfolgeneffekte der Testverfahren wurden durch sechs unterschiedliche Testfolgen kontrolliert (Abbildung 2). Dabei mussten zwei 20- bzw. 30-minütige Zeitintervalle bis zu den verzögerten Abrufbedingung des *California Verbal Learning Tests (CVLT)* und des *Word Memory Tests (WMT)* berücksichtigt werden. Die Fülltests durften wiederum keine mnestischen Funktionen beanspruchen.

Testfolge 1	Testfolge 2	Testfolge 3	Testfolge 4	Testfolge 5	Testfolge 6
WMT Lernen TAP WMT Abruf	WMT Lernen TAP WMT Abruf	AST CVLT Lernen ZVT	AST RCFT WMT Lernen	AKGT CVLT Lernen ZVT	AKGT RCFT WMT Lernen
Instruktionserinnerung 1					
AKGT CVLT Lernen ZVT TL-D	AST CVLT Lernen ZVT TL-D	TL-D CVLT Abruf AKGT RCFT	TAP WMT Abruf CVLT Lernen ZVT	TL-D CVLT Abruf AST RCFT	TAP WMT Abruf CVLT Lernen ZVT
Instruktionserinnerung 2					
CVLT Abruf RCFT AST	CVLT Abruf RCFT AKGT	WMT Lernen TAP WMT Abruf	TL-D CVLT Abruf AKGT	WMT Lernen TAP WMT Abruf	TL-D CVLT Abruf AST

Abbildung 2. Darstellung der sechs Testfolgen

Nach Beendigung der psychometrischen Untersuchung erfolgten ein schriftlicher Manipulationscheck und die Erfassung der Kontrollvariablen sowie der subjektiv

wahrgenommenen Schwierigkeit der einzelnen Testverfahren. Abschließend wurden die Probanden gebeten, separat für jedes Verfahren einzuschätzen, inwieweit sie sich je nach Untersuchungsbedingung entweder bei der Aufgabenbearbeitung angestrengt oder simuliert hatten. Für jede Untersuchungsbedingung wurde dazu ein eigener Fragebogen konstruiert (Anhang I, Anhang J). Eine kurze mündliche Ergebnisrückmeldung erfolgte auf Wunsch fast aller Probanden der KG bereits direkt im Anschluss an die Untersuchung. Am Ende wurde allen Probanden für die Studienteilnahme durch eine kleine Aufmerksamkeit (Schokolade, Pralinen) gedankt, was den Probanden zuvor nicht mitgeteilt worden war.

4.1.1 Instruktionen und Erinnerungsankündigung

Die Instruktionen beider Bedingungen enthielten die gleichen Informationen über die Objektivierung kognitiver Defizite anhand neuropsychologischer Testverfahren im klinischen Alltag. Beiden Gruppen wurde mitgeteilt, dass ihre Testleistung einen Beitrag zur Ermittlung von Referenzwerten für die Diagnostik von Simulationstendenzen in der klinischen Praxis leisten würde. Die Kontrollgruppe (KG) wurde zur maximalen Anstrengungsbereitschaft instruiert (Anhang F). Die experimentellen Simulanten (ES) wurden aufgefordert, kognitive Defizite zu simulieren. Zur Förderung des Rollenverständnisses wurde die Sichtweise eines Simulanten durch eine Fallvignette eines Unfallszenarios veranschaulicht und ausdrücklich auf die Relevanz der Glaubhaftigkeit der kognitiven Defizite verwiesen (Bernard, 1990). Der Wortlaut der Fallvignette wurde von Brockhaus und Merten (2004, S.885) übernommen. Diese enthielt neben detaillierten Informationen über das Szenario auch spezifische Beschwerden zur Symptompräsentation und die Aussicht auf eine finanzielle Entschädigung (Anhang G).

In der mündlichen Überprüfung des Instruktionsverständnisses zu Beginn der Untersuchung wurden die Probanden zur Wiedergabe der Untersuchungsanforderungen mit eigenen Worten aufgefordert. Es wurde geprüft, ob sich die Probanden der KG darüber bewusst waren, während der gesamten Untersuchung ihre bestmögliche Leistung zu zeigen. Das Instruktionsverständnis der ES galt als gegeben, wenn die Probanden sinngemäß angeben konnten, während der Untersuchung zu simulieren und ein Beispiel nennen konnten. Zudem wurde darauf geachtet, dass die Probanden sich bewusst waren, die Defizite möglichst glaubhaft zu simulieren.

Nach der Überprüfung des Instruktionsverständnisses wurde die Erinnerung an die Instruktionsumsetzung angekündigt. Rogers et al. (1993) betonen die Relevanz der Instruktionsumsetzung in Analogstudien und verweisen auf Goebel (1983) in dessen Analogstudie 10% der Probanden die Instruktionen überhaupt nicht umsetzten, da sie sich

selbst als zu ehrlich oder zu motiviert einschätzten. Etwa 50% der Probanden gaben an, die experimentellen Instruktionen nicht konsequent umgesetzt zu haben. Bei Heaton et al. (1978) gaben 20% der Probanden an, überhaupt nicht simuliert zu haben. Auch in aktuelleren Arbeiten verweisen Autoren wie Merten (2014) und Bianchini et al. (2001) auf eine Gefährdung des experimentellen Vorgehens in Analogstudien, da nicht alle Probanden die Instruktionen tatsächlich umsetzen. Demzufolge erhielten die Probanden der ES in der vorliegenden Untersuchung folgende Anleitung:

„Ich werde Sie im Verlauf der Untersuchung zu zwei festgesetzten Zeitpunkten an die Umsetzung der Instruktion erinnern, dass Sie eine kognitiv beeinträchtigte Person simulieren und nicht Ihre tatsächliche Leistung zeigen sollen. Dies erfolgt auf Grund des Aufbaus der Studie und ist unabhängig von ihrer bisherigen Testleistung.“

Analog lautete die Instruktion der KG:

„Ich werde Sie im Verlauf der Untersuchung zu zwei festgesetzten Zeitpunkten an die Umsetzung der Instruktion erinnern, dass Sie sich maximal anstrengen sollen. Dies erfolgt auf Grund des Aufbaus der Studie und ist unabhängig von ihrer bisherigen Testleistung.“

4.1.2 Erhebung psychosozialer und relevanter experimenteller Einflussvariablen

Zu Beginn der Untersuchung wurden auch die psychosozialen Variablen anhand eines kurzen selbstentwickelten Anamnesebogens erfasst (Anhang E). Neben Geburtsdatum, Muttersprache, Familienstand und Beruf wurde der Bildungsabschluss gemäß *CASMIN* (*Comparative Analysis of Social Mobility in Industrial Nations*) nach Lechert, Schroedter und Lüttinger (2006) erfragt. Die CASMIN-Klassifikation wurde in den 1980er Jahren zur Einstufung der sozialen Mobilität in diversen Industrienationen entwickelt. Die international vergleichbare Bildungsklassifikation differenziert das deutsche Bildungssystem in acht Kategorien: Hochschulabschluss, Fachhochschulabschluss, Fachhochschulreife/Abitur und Berufsausbildung, Fachhochschulreife/Abitur ohne Berufsausbildung, Mittlere Reife und Berufsausbildung, Mittlere Reife ohne Berufsausbildung, Hauptschulabschluss, kein Abschluss. In der vorliegenden Arbeit wurde die CASMIN-Klassifikation um die drei Kategorien Fachhochschulreife/Abitur in beruflicher Ausbildung, Mittlere Reife in beruflicher Ausbildung und Hauptschule in beruflicher Ausbildung ergänzt. Darüber hinaus wurde das Vorliegen psychischer und somatischer Erkrankungen sowie die Einnahme verschreibungspflichtiger Medikamente erfasst.

Nach den psychometrischen Testverfahren wurden weitere Einflussvariablen schriftlich in einem postexperimentellen Fragebogen erhoben. Für beide Bedingungen (Instruktionen als

KG oder ES) wurde jeweils eine separate Version entworfen. In beiden Versionen wurden potentielle Einflussvariablen wie Müdigkeit, Langeweile, Interesse und Misstrauen gegenüber der Untersuchung anhand visueller Analogskalen (VAS) gemessen. Die Analogskalen bestanden aus einem 100 mm langen Zahlenstrahl mit Extremausprägungen an den Endpunkten. Während die Endpunkte null (0 = minimal, im Sinne von gar nicht) und hundert (100 = maximal, im Sinne des stärksten Ausprägungsgrades) durch die Begrenzung der Linie vorgegeben sind, finden sich keine Abstufungen auf der Skala. Die Probanden gaben durch ein Kreuz auf der Linie den subjektiv empfundenen Ausprägungsgrad der Variable an. Auch eine Einschätzung der subjektiven Aufgabenschwierigkeit der einzelnen Verfahren erfolgte analog zu Merten et al. (2004) mit Likert-Skalen in beiden Versionen. Der subjektiv wahrgenommene Schwierigkeitsgrad der einzelnen Testverfahren wurde anhand der fünf Abstufungen (0) *sehr einfach*, (1) *einfach*, (2) *mittelschwer*, (3) *schwer* und (4) *sehr schwer* beurteilt.

Der Fragebogen der KG (Anhang I) beinhaltete zudem die Einschätzung der Motivation und der Anstrengungsbereitschaft mittels Analogskalen. Die Beurteilung der subjektiven Schwierigkeit und die Einschätzung der Anstrengungsbereitschaft für jedes Testverfahren erfolgte ebenfalls mit fünfstufigen Likert-Skalen (0) *kaum vorhanden*, (1) *gering*, (2) *mittelgradig*, (3) *stark ausgeprägt* und (4) *sehr stark ausgeprägt*.

Die Fragebogenversion der ES (Anhang J) enthielt neben der Beurteilung potentieller Einflussvariablen und der subjektiven Aufgabenschwierigkeit auch einen schriftlichen Manipulationscheck. In Anlehnung an Bernard (1990) wurden die ES zur sinngemäßen Wiedergabe des Untersuchungsziels aufgefordert, was allen Probanden gelang. Da diverse Autoren die Überprüfung der Rollenkonformität betont haben (Bender & Rogers, 2004; Merten et al., 2004; Nies & Sweet, 1994; Rogers, 1997), erfolgte eine Einschätzung, inwieweit die ES die Instruktionen umgesetzt hatten. Mittels Analogskalen wurde die Simulationsbereitschaft eingeschätzt und sich in einer schlechteren Aufgabenbearbeitung, längeren Bearbeitungszeiten, einer ausgeprägten Vergessens- und Fehlerrate sowie instruktionswidrigem Verhalten und Knapp-daneben-Antworten gezeigt hatte. Analog zur Beurteilung der eigenen Anstrengungsbereitschaft der KG schätzten die ES ihre Simulationsbereitschaft und die subjektive Aufgabenschwierigkeit mittels fünfstufiger Likert-Skalen ein.

4.2 Untersuchungsverfahren

4.2.1 Neuropsychologische Leistungstests

Das kognitive Leistungsniveau wurde mit standardisierten neuropsychologischen Testverfahren erfasst, die im deutschsprachigen Raum häufig Verwendung finden. Die Verfahren wurden so ausgewählt, dass ein breites Spektrum neuropsychologischer Funktionsbereiche und eingebetteter BV-Parameter erfasst werden konnte. Neben der verbalen Lern- und Merkfähigkeit wurden auch basale und komplexe Aufmerksamkeitsfunktionen geprüft. Die visuokonstruktiven und figural-mnestischen Fähigkeiten, die kognitive Informationsverarbeitungsgeschwindigkeit und die exekutiven Funktionen wurden als zusätzliche Leistungstests zusammengefasst.

California Verbal Learning Test (CVLT: Niemann et al., 2008)

Der *California Verbal Learning Test (CVLT)* wurde 1987 von Delis, Kramer, Kaplan und Ober als englischsprachige Forschungsversion zur Erfassung der verbalen Lern- und Merkfähigkeit veröffentlicht. Auf die zweite Version des *CVLT-II* (Delis et al., 2000) folgte eine erneute Überarbeitung, einschließlich einer deutschen Adaption des *CVLT* (Niemann et al., 2008), die in der vorliegenden Arbeit verwendet wurde. Der *CVLT* gilt aktuell als eines der am häufigsten eingesetzten Verfahren zur Beurteilung von Lernen und Gedächtnis (Rabin et al., 2016).

Durchführung: Der *CVLT* setzt sich aus zwei Wortlisten (Liste A und Liste B) zusammen, die jeweils 16 Worte aus vier semantischen Kategorien beinhalten. Es folgen stets Worte unterschiedlicher Kategorien direkt aufeinander. Nachdem der Untersuchungsleiter Liste A laut vorgelesen hat, soll Liste A unmittelbar vom Probanden möglichst vollständig und in beliebiger Reihenfolge repliziert werden (freier Abruf). Dies wird in insgesamt fünf Lerndurchgängen wiederholt, dabei bleibt die Reihenfolge der Stimuluswörter unverändert. Anschließend wird eine Interferenzliste (Liste B) einmal laut vorgelesen und unmittelbar vom Probanden frei erinnert. Dann erfolgen der kurzfristige freie und gestützte Abruf der Liste A. Letzterer beinhaltet die Vorgabe der vier semantischen Kategorien als Abrufhilfe. Nach einem 20-minütigem Behaltensintervall, ohne interferierende verbal-mnestische Aufgaben, werden die verzögerten freien und gestützten Abrufleistungen erfasst. Abschließend wird die Rekognitionsleistung ermittelt. Dabei werden insgesamt 44 Worte (Liste A sowie Worte der Liste B und neue, unbekannte Worte) vorgelesen. Der Proband soll bei jedem der 16 Worte der Liste A mit der Antwort Ja das Wiedererkennen der Worte der Liste A bekunden. Insgesamt beträgt die reine Durchführungszeit ohne das Behaltensintervall ca. 25 Minuten.

Auswertung: Neben einer manuellen Auswertung des *CVLT* ermöglicht die differenziertere computergestützte Auswertung 28 bildungs- und altersnormierte Leistungsparameter. In die vorliegende Arbeit wurden die in der Praxis häufig verwendeten 12 Kennwerte der manuellen Auswertung einbezogen. 1) Anzahl korrekt erinnertes Wörter aus Liste A nach dem ersten Lerndurchgang (*Durchgang 1: DG1*), 2) Anzahl korrekt erinnertes Wörter aus Liste A nach dem fünften Lerndurchgang (*Durchgang 5: DG5*), 3) Lernsumme als Gesamtanzahl korrekt erinnertes Wörter über fünf Lerndurchgänge (*Lernsumme: LS*), 4) Anzahl korrekt erinnertes Wörter der *Interferenzliste B* (*Interferenzliste B: B*), 5) Anzahl korrekt erinnertes Wörter nach kurzem Zeitintervall ohne semantische Hinweisreize (*Verzögerte Freie Wiedergabe I: VFW I*), 6) Anzahl korrekt erinnertes Wörter nach kurzem Zeitintervall mit semantischen Hinweisreizen (*Wiedergabe mit Abrufhilfe I: WA I*), 7) Anzahl korrekt erinnertes Wörter nach 20-minütigem verzögerten Zeitintervall ohne semantische Hinweisreize (*Verzögerte Freie Wiedergabe II: VFW II*), 8) Anzahl korrekt erinnertes Wörter nach 20-minütigem Verzögerungsintervall mit semantischen Hinweisreizen (*Wiedergabe mit Abrufhilfe II: WA II*), 9) Summe der Intrusionen während der Wiedergabe mit Abrufhilfe (*Intrusionen Wiedergabe mit Abrufhilfe: IWA*), 10) Anzahl richtig erinnertes Antworten bei der Rekognitionsprüfung (*Trefferrate: T*), 11) Anzahl fälschlicherweise erinnertes Wörter bei der Rekognitionsprüfung (*Falsch-Positive: FP*), 12) Unterscheidungsfähigkeit zwischen Zielitems und Distraktoren, Diskriminationsleistung als prozentualer Anteil richtig klassifizierter Wörter bei der Rekognitionsprüfung mit einer parametrischen Auswertung (*Diskriminabilität d': d'*).

Gütekriterien: Die Retest-Reliabilitätskoeffizienten rangieren zwischen $r_{tt} = .26$ (*Interferenzliste B*) und $r_{tt} = .81$ (*Wiedergabe mit Abrufhilfe II: WA II*, *Trefferrate T*). Die Konstruktvalidität gilt durch sechs ermittelte Faktoren (Allgemeine verbale Lern- und Abruffähigkeit, Diskriminationsfähigkeit, Serieller Positionseffekt, serielle und semantische Lernstrategie, Lernverlauf und Wiedererkennen, Proaktive Interferenz) als bestätigt, die insgesamt bis zu 73% der Gesamtvarianz aufklären.

Testbatterie zur Aufmerksamkeitsprüfung (TAP: Zimmermann & Fimm, 2012)

Die erste Version der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* wurde 1992 von Zimmermann und Fimm veröffentlicht. Die *TAP* erfasst computergestützt spezifische Aspekte attentionaler und exekutiver Funktionsbereiche. In dieser Studie wurden die drei Untertests *Alertness*, *geteilte Aufmerksamkeit* und *Go/Nogo* der insgesamt 13 Einzeltests der Version 2.3 verwendet (Zimmermann & Fimm, 2012). Die visuelle Stimuluspräsentation erfolgt am Bildschirm, das auditive Reizmaterial wird mittels Lautsprecher präsentiert. Die

Reaktionszeiten werden durch den Tastendruck auf eine Reaktionszeit-Taste gemessen. Vor jedem Untertest wird ein separater Vortest zur Überprüfung des Instruktionsverständnisses durchgeführt, der bei der Auswertung nicht berücksichtigt wird.

Untertest Alertness

Als *Alertness* bezeichnet man den allgemeinen Wachzustand als Grundvoraussetzung aller Aufmerksamkeitsleistungen. Dadurch werden schnelle und adäquate Reaktionen und situationsspezifische Verhaltensweisen erst ermöglicht. Man unterscheidet zwischen dem allgemeinen Wachzustand (*tonische Arousal*), einschließlich dessen Tagesschwankungen, und einem kurzfristigen Fokussieren der Aufmerksamkeit in Erwartung eines Ereignisses (*phasische Arousal*). Ein Beispiel aus dem Alltag stellt die Reaktionsbereitschaft dar, die beim raschen Bremsen und Anfahren im Stop and Go-Verkehr beansprucht wird.

Durchführung: Der Untertest *Alertness* besteht aus zwei Bedingungen. Der Proband wird instruiert, bei Erscheinen eines Kreuzes so schnell wie möglich die Reaktionszeit-Taste zu drücken. Die Darstellung des Kreuzes in der Bildschirmmitte bleibt dabei konstant, die Zeitintervalle des Erscheinens des Kreuzes variieren zufällig. Bei Bedingung A wird nur das Kreuz eingeblendet, während Bedingung B zusätzlich einen Warnton enthält. Dieser ertönt jeweils vor dem Erscheinen des Kreuzes. In beiden Bedingungen soll der Proband bei Erscheinen des Kreuzes die Reaktionszeit-Taste drücken. Zur Vermeidung von Ermüdungseffekten sind die beiden Bedingungen (Präsentation von jeweils 20 Zielreizen) als ABBA-Design angelegt.

Die reine Durchführungszeit beträgt ohne Instruktion und Vortest 4 Minuten 30 Sekunden.

Auswertung: Es werden jeweils für die gesamte Testdurchführung, für jeden einzelnen Durchgang und für die Bedingungen mit und ohne Warnton der Mittelwert, der Median und die Standardabweichungen der Reaktionszeiten sowie jeweils die Anzahl der richtigen Reaktionen, der Reaktionsauslassungen, der Ausreißer (Reaktionszeiten größer als die mittlere Reaktionszeiten plus 2.35 mal die Standardabweichungen der Reaktionszeiten) und der Antizipationen (Reaktionen auf den Warnton) ausgegeben. Darüber hinaus wird ein Kennwert der phasischen Alertness ermittelt, der die Fähigkeit abbildet, in Erwartung eines kritischen Reizes die Aufmerksamkeit zu erhöhen. Sie wird durch die Verkürzung der Reaktionszeit auf einen angekündigten im Vergleich zu einem unangekündigten Reiz untersucht. Die Normwerte für Erwachsene wurden aus einer Stichprobe mit 604 Personen (Alter 19 bis 89 Jahre) gewonnen. In die vorliegende Arbeit wurden alle Parameter mit verfügbaren

Normwerten eingeschlossen: *Reaktionszeit-Mediane mit Warnton (Am Md)* und *ohne Warnton (Ao Md)*, *Standardabweichung der Reaktionszeiten mit Warnton (Am SD)* und *ohne Warnton (Ao SD)*, *Kennwert phasische Alertness (Apha)*.

Gütekriterien: Die Retest-Reliabilität beträgt bei einem durchschnittlichen Zeitintervall von 25 Tagen $r_{tt} = .81$ für den *Median der Reaktionszeiten* des Gesamttestes.

Untertest Geteilte Aufmerksamkeit

Unsere Fähigkeit, die Aufmerksamkeit zu teilen, um gleichzeitig ablaufende Prozesse verarbeiten zu können, stellt eine notwendige Funktion zur Alltagsbewältigung dar. Während die Relevanz dieser komplexen Aufmerksamkeitsfunktion als unumstritten gilt, liegt eine Kontroverse der theoretischen Annahmen vor. Neben einer Kapazitätstheorie, die eine Aufteilung und Steigerung kognitiver Ressourcen durch Anstrengung bei einer simultanen Aufgabenbearbeitung postuliert (Posner & Rafal, 1987), wird auch eine *Shifting*-Theorie (Allport, 1993) vertreten. Gemäß der *Shifting*-Theorie wird eine simultane Aufgabenbearbeitung nur durch den Wechsel des Aufmerksamkeitsfokus zwischen den konkurrierenden Anforderungen ermöglicht.

Durchführung: In der vorliegenden Studie wurde die Doppelaufgabe der auditiv-visuellen Bedingung der Durchführungsform I (synchron) verwendet. Diese setzt sich aus zwei unterschiedlichen Aufgaben (visuell und auditiv) zusammen, die gleichzeitig bearbeitet werden sollen. Die visuelle Aufgabe besteht aus einem 4x4-Punkte-Feld, das in der Bildschirmmitte eingeblendet wird und 16 Positionen markiert. In einem gleichförmigen Rhythmus erscheinen bis zu acht Kreuze (x), die unterschiedliche Muster bilden. Sobald vier Kreuze an direkt anliegenden Positionen aufleuchten und die Ecken eines kleinen Quadrats darstellen, soll der Proband so schnell wie möglich die Reaktionszeit-Taste drücken. Gleichzeitig bearbeitet der Proband eine auditive Aufgabe. In einem synchronen Rhythmus zu den Kreuzen werden alternierend ein hoher und ein tiefer Ton eingespielt. Sobald der gleiche Ton zweimal hintereinander zu hören ist, soll der Proband ebenfalls so schnell wie möglich die Reaktionszeit-Taste betätigen.

Die Durchführungszeit beträgt ohne Instruktion und Vortest 3 Minuten 25 Sekunden.

Auswertung: Für die visuellen und auditiven Durchgänge werden jeweils folgende Leistungsparameter berechnet: Mittelwert, Median, Standardabweichung der Reaktionszeiten, Anzahl der korrekten Reaktionen, Reaktionsauslassungen und Ausreißer. Für den Gesamttest wird die Anzahl der Reaktionsauslassungen und Fehlreaktionen ausgegeben. Es liegen Normwerte für den *Median der Reaktionszeit (GAa Md, GAv Md)*, die *Standardabweichung*

der Reaktionszeit ($GAa SD$, $GAv SD$), die Anzahl an Reaktionsauslassungen ($GAa A$, $GAv A$) visueller und auditiver Durchgänge und für die Gesamtzahl der Reaktionsauslassungen ($GA A$) und Fehlreaktionen ($GA F$) vor, die allesamt in die Studie einbezogen wurden.

Gütekriterien: Die Koeffizienten der Retest-Reliabilität liegen bei einem durchschnittlichen Zeitintervall von 25 Tagen zwischen den beiden Messungen bei $r_{tt} = .48$ für den Median der Reaktionszeiten, $r_{tt} = .64$ für die Fehlreaktionen und $r_{tt} = .44$ für die Reaktionsauslassungen.

Untertest Go/Nogo

Mit Hilfe des *Go/Nogo*-Paradigmas wird die Fähigkeit zur Verhaltenskontrolle erfasst. Dabei wird unter Zeitdruck ein unangemessener Verhaltensimpuls auf einen externen Stimulus unterdrückt, um stattdessen eine intern kontrollierte Reaktion ausführen zu können. Der Aufmerksamkeitsfokus wird auf den erwarteten Stimulus gerichtet, dessen Eintreten eine selektive Reaktion (Reaktionsausführung oder Reaktionsunterdrückung) erfordert.

Durchführung: In der vorliegenden Arbeit wurde die Durchführungsform „1 aus 2“ (ein kritischer Reiz von zwei Reizen) verwendet, in der zwei gut diskriminierbare Stimuli dargeboten werden. In der Mitte des Bildschirms erscheinen in wechselnder Abfolge entweder ein liegendes Kreuz (x = kritischer Reiz) oder ein stehendes Kreuz ($+$ = nicht-kritischer Reiz erfordert Reaktionsinhibition). Durch die kurze Darbietungszeit (200 ms) der jeweils 20 kritischen und nicht-kritischen Reize wird ein unmittelbarer Reaktionsimpuls ausgelöst. Die Aufgabe besteht darin, bei Erscheinen des kritischen Reizes (x) so schnell wie möglich die Reaktionszeit-Taste zu drücken und beim nicht-kritischer Reiz ($+$) nicht zu reagieren.

Die Durchführungszeit (ohne Vortest und Instruktion) beträgt zwei Minuten.

Auswertung: Für den Gesamttest werden der Mittelwert, der Median und die Standardabweichung der Reaktionszeiten, die Anzahl der korrekten Reaktionen, der Fehlreaktionen, der Reaktionsauslassungen sowie der Ausreißer angegeben. Normwerte liegen für den Median der Reaktionszeiten ($GN Md$), die Standardabweichung der Reaktionszeiten ($GN SD$), die Anzahl Fehlreaktionen ($GN F$) und Reaktionsauslassungen ($GN A$) vor und wurden in die folgenden Analysen aufgenommen.

Gütekriterien: Die Koeffizienten der Retest-Reliabilität betragen $r_{tt} = .56$ für den Median der Reaktionszeiten, $r_{tt} = .73$ für die Fehlreaktionen und $r_{tt} = -.09$ für die Reaktionsauslassungen.

4.2.2 **Additionalere Leistungstests**

Rey Complex Figure Test and Recognition Trial (RCFT: Meyers & Meyers, 1995)

Der *Rey Complex Figure Test and Recognition Trial (RCFT)* ist ein Paper-Pencil-Verfahren zur Erfassung visuokonstruktiver Fähigkeiten und visueller Gedächtnisleistungen (Meyers & Meyers, 1995). Das Verfahren enthält die komplexe geometrische Rey-Figur (Rey, 1941), eine Weiterentwicklung der Rey-Figur von Osterrieth (Osterrieth, 1944).

Durchführung: Der *RCFT* besteht aus vier Durchgängen. Zunächst soll die detailreiche Rey-Figur als Testvorlage so genau wie möglich ohne die Verwendung von Hilfsmitteln abgezeichnet werden (*Kopie*). Nach einem dreiminütigen (*kurz verzögerter Abruf*) und einem 30-minütigen Zeitintervall (*lang verzögerter Abruf*) wird der Proband jeweils zur freien Reproduktion dieser Figur ohne Zeitdruck aufgefordert. Abschließend erfolgt das Wiedererkennen einzelner geometrischer Elemente der Figur. In der vorliegenden Arbeit wurde auf die beiden letzten Durchgänge auf Grund des 30-minütigen Pausenintervalls verzichtet.

Auswertung: Zur Auswertung des *RCFT* werden 18 Segmente der Rey-Figur herangezogen (Corwin & Bylsma, 1993; Lezak, Howieson & Loring, 1995; Rey, 1941). Jedes Segment wird hinsichtlich Genauigkeit und Platzierung bewertet. Insgesamt rangieren die Rohwertpunkte des *RCFT* zwischen 0 und 36, die anhand altersnormierter Referenzwerte in Prozentränge und *T*-Werte transformiert werden können.

Gütekriterien: Die Koeffizienten Interrater-Reliabilität variieren zwischen $r = .93$ und $r = .99$ (Median $r = .94$). Die Konstruktvalidität mit den *Wechsler Intelligenzskalen* (Wechsler, 1981) liegt zwischen $r = .30$ und $r = .59$.

Turm von London - Deutsche Version (TL-D: Tucha & Lange, 2004)

Der *Turm von London (TL-D)* wurde zur Erfassung des konvergenten, räumlich-problemlösenden Denkens entwickelt (Shallice, 1982) und setzte sich, analog zum *Turm von Hanoi* (Simon, 1975), aus Transformationsaufgaben zusammen.

Durchführung: Der *TL-D* besteht aus drei verschiedenfarbigen Holzkugeln, die auf drei senkrecht nebeneinander angeordnete Holzstäbe gesetzt werden. Die Stäbe sind unterschiedlich lang, so dass entweder eine, zwei oder drei Kugeln auf die Stangen passen. Die Kugeln sollen von einem Ausgangszustand durch eine bestimmte Anzahl an Zügen in einen vorgegebenen Zielzustand transformiert werden. Der Schwierigkeitsgrad nimmt mit

steigender Anzahl vorgegebener Züge zu. Es liegen insgesamt 20 Aufgaben unterschiedlichen Schweregrads vor (jeweils fünf 3-Zug-, 4-Zug-, 5-Zug- und 6-Zug-Probleme).

Die Durchführungszeit beträgt etwa 20 bis 25 Minuten.

Auswertung: Die Bearbeitungszeit wird dokumentiert, die Testleistung ergibt sich jedoch ausschließlich aus der Gesamtzahl der Aufgaben, die mit der vorgegebenen Anzahl an Zügen gelöst wurden. Die Auswertung erfolgt mit Normwerten, die nach Bildungsjahren und Alter untergliedert sind. Dem Rohwert der Gesamtleistung wird so der entsprechende Prozentrang zugeordnet.

Gütekriterien: Die Retest-Reliabilitätskoeffizienten variieren zwischen $r_{tt} = 1.0$ und $r_{tt} = .85$ für die verschiedenen Aufgabenschwierigkeiten. Der Koeffizient der Split-Half-Reliabilität liegt bei $r = .86$ ($N = 630$ gesunde Erwachsene). Die interne Konsistenz über alle 20 Aufgaben beträgt $\alpha = .76$. Die Konstruktvalidität rangiert zwischen $r = -.54$ (kognitive Flexibilität) und $r = .30$ (verbale Merkspanne).

Zahlen-Verbindungs-Test (ZVT: Oswald & Roth, 1987)

Der *Zahlen-Verbindungs-Test (ZVT)* erfasst die zentralnervöse Informationsverarbeitungsgeschwindigkeit. In der Literatur wird diese basale, altersabhängige Intelligenzfunktion häufig mit fluiden Intelligenzaspekten gleichgesetzt.

Durchführung: In der vorliegenden Arbeit wurde die 2. überarbeitete und erweiterte Auflage des *ZVT* (Oswald & Roth, 1987) verwendet. Das Untersuchungsmaterial des Paper-Pencil-Tests besteht aus zwei verkürzten Übungsmatrizen sowie vier Zahlen-Matrizen mit jeweils 90 unterschiedlich angeordneten Ziffern. Die Ziffern sollen in aufsteigender, numerischer Reihenfolge so schnell wie möglich miteinander verbunden werden.

Die Durchführungszeit variiert zwischen 5 und 10 Minuten.

Auswertung: Aus der Summe der Bearbeitungszeiten der vier Matrizen wird der Mittelwert der Durchführungszeit in Sekunden berechnet. Anhand altersgestaffelter Normwerte (8 bis 60 Jahre, $N = 2109$) kann der Mittelwert in einen *T*-, *c*- und *PR*-Wert sowie in einen IQ- und SW-Wert transformiert werden.

Gütekriterien: Die Korrelationskoeffizienten der Konstruktvalidität rangieren zwischen $r = -.40$ mit den *Standard Progressive Matrices* (Raven, 1983) und $r = -.83$ mit dem *Prüfsystem für Schul- und Bildungsberatung* (Horn, 1969). Kurze Bearbeitungszeiten im *ZVT* korrelieren mit überdurchschnittlichen Leistungsparametern der Intelligenzverfahren.

4.2.3 Performanzvalidierungstests (PVT)

Word Memory Test (Green, 2003, Revised 2005)

Der *Word Memory Test (WMT)* wurde 1994 von Paul Green als erstes Verfahren entwickelt, das neben der Gedächtnisleistung auch die Anstrengungsbereitschaft einer Person erfassen kann. Die englischsprachige Originalform lag als mündliche Testversion vor (Green, Allen & Astner, 1996), deren Aufbau und Inhalte in der computergestützten Version beibehalten wurden. In der vorliegenden Arbeit wurde die deutschsprachige Papier-Pencil-Version des *WMT* (Green, 2003, Revised 2005) zur Erfassung der Anstrengungsbereitschaft verwendet, auf die Beurteilung der Gedächtnisleistung anhand des *WMT* wurde verzichtet.

Durchführung: Der *WMT* besteht aus einer Lernphase und einer anschließenden Testphase, die sich aus sechs Subtests zusammensetzt. Während der Lernphase wird eine Wortliste bestehend aus 20 Wortpaaren (z.B. Hund-Katze) zweimal hintereinander präsentiert, dabei bleiben die Wortpaare und deren Reihenfolge unverändert. Durch die Anzahl von 40 Stimuluswörtern wird eine höhere subjektive Testschwierigkeit induziert, als diese tatsächlich vorliegt. Die anschließende Testphase beginnt mit dem *unmittelbaren Wiedererkennenstest (Immediate Recognition, IR)*, in dem 40 neue Wortpaare vorgelesen werden. Jedes dieser Wortpaare setzt sich aus einem Wort der Originalliste und einem neuen Distraktorwort zusammen, wobei das aus der Originalliste stammende Wort zu erkennen ist. Nach einem 30-minütigen Zeitintervall erfolgt der *verzögerte Wiedererkennenstest (Delayed Recognition, DR)* der Wörter der Originalliste mittels einer weiteren Wortliste. Aus diesen beiden Untertests wird die Einschätzung der Anstrengungsbereitschaft ermittelt.

Auswertung und Interpretation: Aus der Anzahl richtig erkannter Wörter wird im Vergleich zur Gesamtzahl von 40 Wörtern für das *unmittelbare Wiedererkennen (IR)* und das *verzögerte Wiedererkennen (DR)* jeweils ein Prozentwert (*Pir* und *Pdr*) ermittelt. Der Vergleich der Wiedererkennensleistung zwischen den beiden Durchgängen ermöglicht die Berechnung eines *Konsistenzmaßes (Con)*. Somit ergeben sich aus der Gedächtnisleistung drei Kennwerte zur Beurteilung der Anstrengungsbereitschaft (*IR, DR, Con*), die unabhängig vom Alter, Ausbildungsgrad und der verbalen Intelligenzleistungen der jeweiligen Person sind. Bei Prozentwerten von mehr als 90% wird die Anstrengungsbereitschaft als unauffällig interpretiert. Der Cut-off-Wert für suboptimale Anstrengungsbereitschaft liegt bei 82.5%. Dieser Wert liegt für das *unmittelbare Wiedererkennen (IR)* 5.5 *SD* und für das *verzögerte Wiedererkennen (DR)* 6.7 *SD* unterhalb der Kontrollgruppe gesunder Erwachsener und 3 *SD* unterhalb der Vergleichsgruppe mit kognitiv-neurologischen Patienten. Bereits ab einem Wert

zwischen 83% und 90% sind die kognitiven Leistungsparameter einer Person mit Vorbehalt zu betrachten.

Die Subtests *unmittelbares Wiedererkennen (IR)* und *verzögertes Wiedererkennen (DR)* weisen den gleichen Schwierigkeitsgrad auf. Somit deuten unterschiedliche Leistungen in den beiden Tests, wie auch ein geringes *Konsistenzmaß (Con)*, auf eine verminderte Reliabilität und Validität der Testergebnisse. Ein auffälliges Ergebnis in einem Subtest ist demnach nicht als Anzeichen einer unverminderten Anstrengungsbereitschaft fehlzuinterpretieren, sondern gilt als weiterer Hinweis für eine suboptimale Anstrengungsbereitschaft. Im Testhandbuch liegen keine Angaben über den Wertebereich der Ratewahrscheinlichkeit vor. Dieses Intervall (34.5% bis 65.5%) wurde mit dem Standardfehler der Ratewahrscheinlichkeit von $p = .50$ unter Einbezug der Irrtumswahrscheinlichkeit von $\alpha = .05$ und der Standardnormalverteilung ergänzt. Zur Berechnung wurde die Freeware *Berechnung des Konfidenzintervalls eines Anteilswerts* der Universität Köln herangezogen (Andreß, 2003).

Die Gedächtnisleistung kann qualitativ (bessere Leistungen in einfacheren als in schwierigeren Subtests) und quantitativ durch eine z -Transformation ausgewertet werden.

In Bezug auf die Durchführungsdauer sind im Testmanual keine Angaben zu finden.

Gütekriterien: Im Testhandbuch wird eine Within-Test-Reliabilität ($N = 1207$) der Parameter *Pir* und *Pdr* $r = 0.88$ aufgeführt. Der Korrelationskoeffizient der Retest-Reliabilität (*Pir*, *Pdr* und *Pcon*) lag in einer Analogstudie mit 40 gesunden Erwachsenen bei $r_{tt} = 0.97$. Aus zwei Analogstudien mit experimentellen Simulanten ergaben sich eine Sensitivität von 98% und eine Spezifität von 100%.

Amsterdamer Kurzzeitgedächtnistest (AKGT: Schmand et al., 2005b)

2005 wurde neben der niederländischen Originalversion auch der deutschsprachige *Amsterdamer Kurzzeitgedächtnistest (AKGT: Schmand et al., 2005b)* und der englischsprachige *Amsterdam Short-Term Memory Test (ASTM: Schmand et al., 2005a)* herausgegeben. Der *AKGT* wird den Probanden als Gedächtnis- und Konzentrationstests präsentiert. Durch die Mehrfachwahl von drei Richtigen unter fünf Alternativen, stellt der *AKGT* eine Modifikation der Alternativwahlverfahren dar (Merten, 2014). Dadurch soll die eigentliche Testabsicht und das zugrunde liegende Testprinzip geschützt werden.

Durchführung: Das computergestützte Verfahren setzt sich aus zwei Übungsdurchgängen und 30 Testaufgaben zusammen. Jedes Testitem besteht aus drei Bildschirmseiten. Auf der ersten Seite wird eine Liste mit fünf Stimuluswörtern einer bestimmten Kategorie (z.B. Fische) für acht Sekunden eingeblendet, die von dem Probanden laut vorgelesen und eingepägt werden

soll. Jede Wortliste beinhaltet die fünf häufigsten Vertreter der einzelnen Kategorien. Wiederrum drei Wörter der Wortlisten sollen jeweils nach einer einfachen mathematischen Distraktoraufgabe (Addition oder Subtraktion) aus einer anderen Wortliste mit zwei weiteren Wörtern (seltene Kategorievertreter) wiedererkannt und dem Testleiter genannt werden. Der Proband erhält jeweils Rückmeldung über die Richtigkeit seiner Antworten.

Nach Bearbeitung der ersten 15 Testitems besteht die Möglichkeit das Verfahren zu beenden, insofern mehr als sechs Fehler gemacht wurden oder eine zügige Bearbeitung mit maximal zwei Fehlern vorliegt. Die Autoren raten jedoch zur vollständigen Testdurchführung, da Personen mit mangelnder Anstrengungsbereitschaft häufig in der zweiten Testhälfte mehr Fehler zeigen. Bei Rechenschwierigkeiten in den beiden Übungsdurchgängen empfehlen die Autoren auf die Testversion mit vereinfachten Rechenaufgaben zurückzugreifen. In der vorliegenden Analogstudie wurde der *AKGT* mit allen 30 Items vollständig bearbeitet, die Patienten mit authentischen kognitiven Beeinträchtigungen erhielten die verkürzte Version. Die Durchführungszeit beträgt 10 bis 15 Minuten. Die Bearbeitung mit verminderter Anstrengungsbereitschaft kann mit längeren Bearbeitungszeiten einhergehen.

Auswertung und Interpretation: Die Testauswertung erfolgt mit der Summe richtig erkannter Wörter (*AKGT G*). Ein Rohwert ≤ 84 der Maximalpunktzahl von 90 deutet auf eine mangelnde Anstrengungsbereitschaft. Ein Antwortmuster im Bereich der Ratewahrscheinlichkeit resultiert ab einem Wert von 54 Punkten. Summenwerte ≤ 48 Punkte liegen statistisch signifikant unterhalb des Zufallsniveaus. Zur Unterscheidung zwischen dem Gesamtwert *AKGT G* und der Anzahl richtiger Antworten in der ersten Testhälfte wurde für die vorliegende Arbeit der Kennwert *AKGT I* gebildet.

Gütekriterien: Der Retest-Reliabilität-Koeffizient beträgt bei bona-fide Patienten mit leichten bis schweren kognitiven Beeinträchtigungen $r_{tt} = .84$. Die Produkt-Moment-Korrelation (Pearson) zwischen den beiden Testungen liegt bei $r = .91$, die Spearman-Rangkorrelation bei $r = .84$ (jeweils $p < .05$). Der optimale Trennwert von 84/85 Rohwertpunkten weist eine Sensitivität von 91% und eine Spezifität von 89% auf. Laut diverser experimenteller Validierungsstudien des Testmanuals (de Sterke, 1993; Schagen, Schmand, de Sterke & Lindeboom, 1997; Stapert, 1997; Bolan, Foster, Schmand & Bolan, 2002; Merten, Henry & Hilsabeck, 2004; Henry, Balskewitz, Brockhaus, Green & Merten, 2004) liegen zuverlässige Klassifikationsstatistiken vor ($75\% \leq \text{Sensitivität} \leq 100\%$, $89\% \leq \text{Spezifität} \leq 100\%$).

Aggravations- und Simulationstest (AST: Eberl & Wilhelm, 2007)

Die erste Auflage des computergestützten *Aggravations- und Simulationstests (AST)* von Eberl und Wilhelm erschien 2004. In der vorliegenden Arbeit wurde die zweite Auflage des deutschsprachigen AST 4.0 (Eberl & Wilhelm, 2007) verwendet.

Durchführung: Das Testverfahren basiert auf einem Alternativwahlformat mit Gedächtnisaufgaben. Nach zwei Beispielaufgaben mit Ergebnismeldung erfolgt die Bearbeitung der 70 Testitems ohne Rückmeldungen. Dem Probanden wird dabei ein zu memorierender Stimulus für 1.5 Sekunden auf dem Bildschirm präsentiert. Nach einem ebenfalls 1.5-sekündigen Zeitintervall mit leerem Bildschirm soll dieses Zielitem aus zwei Antwortreizen wiedererkannt und per Computertastatur ausgewählt werden. Das Prinzip der verdeckten Leichtigkeit wird durch einen Wechsel des zu memorierenden Stimulusmaterials operationalisiert. Die Stimuli variieren in Form unterschiedlicher Modalitäten durch abstrakte Bilder sowie nicht-abstrakter Items (konkrete Bilder, Wörter und Zahlen). Danach stehen stets zwei Stimuli zur Auswahl. Diese weisen entweder die gleiche Modalität wie das zuvor präsentierte Zielitem auf (kongruent) oder unterscheiden sich vom Zielitem (inkongruent). Daraus ergeben sich sieben Kombinationsmöglichkeiten, die als Objektklassen bezeichnet werden. Jede Objektklasse beinhaltet zehn Items.

Die Durchführungsdauer wird im Testhandbuch nicht angegeben.

Auswertung und Interpretation: Der *AST* gilt als alters-, geschlechts- und bildungsunabhängig. Die computergestützte Auswertung basiert auf einer Normstichprobe aus einer gesunden KG ($n = 51$), ES ($n = 32$) und kognitiv beeinträchtigten bona-fide Patienten ($n = 39$), die sich weder hinsichtlich Geschlecht, Bildung noch Alter unterscheiden. Als „simulationssensible Werte“ nennen Eberl und Wilhelm (2007, S.16) die drei Variablen *Prozent richtiger Antworten (AST P)*, *Median der Reaktionszeiten (AST Md)* und *Standardabweichung der Reaktionszeiten (AST SD)*. In der Validierungsstudie der Autoren unterschieden sich die Gruppen vor allem in der Variablen *AST P*. Bei einem Grenzwert von acht Fehlern bzw. einem *AST P* von 88.57% lagen die Sensitivität und Spezifität jeweils bei 100%. Die Autoren orientierten sich zudem an den Empfehlungen von Guilmette et al. (1993) sowie Hiscock und Hiscock (1989) und definierten einen Grenzwert von 90% richtiger Antworten. Da die Autoren keine Angaben über den Bereich der Ratewahrscheinlichkeit des *AST P* angegeben haben, wurde dieser durch eigene Reanalysen ergänzt (39.67% bis 60.33%). Für die Variablen *AST M* und *AST SD* wurden im Manual keine Grenzwerte angegeben. Interferenzstatistisch zeigten sich allerdings signifikante Gruppenunterschiede. Die ES

erzielten langsamere Reaktionszeiten als die beiden anderen Gruppen, die Patienten zeigten größere Standardabweichungen als die KG und ES.

Gütekriterien: Die interne Konsistenz der Variable *Prozent richtiger Antworten (AST P)* lag bei $\alpha = .97$ für die Gesamtstichprobe, bei $\alpha = .96$ für die ES. Die interne Konsistenz der Variable *Median der Reaktionszeiten (AST Md)* betrug $\alpha = .89$ für die Gesamtstichprobe, $\alpha = .92$ für die ES. Eine hohe Konstruktvalidität begründeten die Autoren durch den bewährten Einsatz von Alternativwahlverfahren. Die Kriteriumsvalidität wurde durch die Korrelation mit anderen kognitiven Leistungsparametern als Außenkriterien definiert. Bei gesunden Probanden korrelierte der *AST P* mit der Konzentrationsleistung der *Digit Span forward* des *Wechsler Gedächtnistest-Revidierte Fassung (WMS-R: Härting et al., 2000)* zu $r = .40$. In der Patientenstichprobe zeigte sich ein positiver Zusammenhang von $r = .39$ zwischen dem *AST P* und dem Arbeitsgedächtnis in der *Digit Span backward* im *WMS-R*.

4.3 Fallzahlberechnung

Die Konzeption der Analogstudie beinhaltete auch eine Stichprobenumfangsplanung und eine a priori Teststärkenbestimmung (Tabelle 7).

In bisherigen Analogstudien wurden sehr große Gruppenunterschiede mit Effektstärken zwischen $d = -3.59$ und $d = -6.33$ im *Amsterdamer Kurzzeitgedächtnistest (AKGT)* und *Word Memory Test (WMT)* ermittelt (Blaskewitz & Merten, 2006; Bolan et al., 2002; Merten et al., 2004). Auf Grund kleinerer Gruppenunterschiede bei eingebetteten Parametern wurden zur Fallzahlbestimmung Studien mit dem *California Verbal Learning Test (CVLT)* und der *Testatterie zur Aufmerksamkeitsprüfung (TAP)* herangezogen. Neben Analogstudien wurden auch *known-group Studien* berücksichtigt. Fehlende Effektstärken in den Originalstudien wurden aus den entsprechenden statistischen Kennwerten (Mittelwert, Standardabweichung, Stichprobengröße) mit dem Programm G*Power 3.1.9.2 ergänzt (Faul, Erdfelder, Buchner & Lang, 2009). Die Effektstärke berechnete sich aus der Differenz der beiden Stichprobenmittelwerte, dividiert durch die geschätzte Populationsstreuung. Die Populationsstreuung ergab sich als gepolte Standardabweichung aus den jeweiligen Streuungen der beiden Mittelwerte (Rasch, Friese, Hofmann & Naumann, 2010). Es wurden ein Signifikanzniveau von $p = .05$ und eine Teststärke ($1-\beta$) von 80% festgelegt. Die Effektstärken der eingebetteten Parameter der *TAP* lagen zwischen $d = 1.70$ und $d = 11.11$. Die Effektstärken im *CVLT-I* rangierten zwischen $d = -0.65$ und $d = -2.39$. Für die Berechnung der Effektstärken des *CVLT-II* konnten nur Studien mit Patientenstichproben in

die Berechnungen einbezogen werden. Die Effektstärken lagen unter den Kennwerten der Analogstudien ($d = -0.06$ und $d = -1.48$). Insgesamt rangierten die Stichprobengrößen zwischen $N = 4$ und $N = 60$. Aus ökonomischen Gründen und der Empfehlung von $n = 30$ einzelner Gruppengrößen zur Verwendung gängiger Teststatistiken wurde eine Stichprobengröße von $N = 66$ Probanden angestrebt.

Tabelle 7
Stichprobenumfangsplanung aus Effektstärken der Forschungsliteratur.

Autoren	Gruppenvergleich (n)	Parameter	Cohens <i>d</i>	<i>N</i> aus <i>d</i>
TAP: Alertness				
Fiene, Bittner,	Kontrollgruppe (17) vs.	AoM	11.11	4
Fischer, et al.	Experimentelle Simulanten (17)	AoSD	9.27	4
(2015)		AoAuslassungen	1.59	12
		AmM	10.00	4
		AmSD	8.54	4
		AmAuslassungen	1.70	18
		AmAntizipationen	4.14	4
CVLT-I				
Demakis (1999)	Kontrollgruppe (21) vs. Experimentelle Simulanten (21)	LS	-1.55	12
		VFW I	-1.99	10
		VFW II	-2.04	8
Demakis (2004)	Kontrollgruppe (27) vs. Experimentelle Simulanten (26)	LS	-1.2 a	20
		VFW II	-1.4 a	16
		T	-1.3 a	18
		d	-1.4 a	16
Sweet, Wolfe,	Kontrollgruppe (21) vs. Experimentelle Simulanten (25)	LS	-1.98	28
Sattlberger et al.		VFW II	-1.94	28
(2000)		T	-2.33	8
		d	-2.39	8
		Intrusionen Rekognition/ Recall	-0.78 -0.65	42 60
CVLT-II				
Whiteside et al.	Patienten mit Schädel-Hirn- Trauma (91) vs. Klinische Simulanten (67)	T	-1.48	14
(2015)		FP	-0.52	94
		DG5	-1.12	22
Wolfe et al. (2010)	Patienten mit Schädel-Hirn- Trauma (125) vs. Klinische Simulanten (77)	VFW II	-0.06	6030
		<i>d'</i>	-0.61	68
		DW	-1.04	26

Anmerkungen. ^a Kennwert aus Originalstudie.

CVLT: *DG5*: Liste A Durchgang 5, *LS*: Liste A Lernsumme, *VFW I*: Verzögerte freie Wiedergabe I, *VFW II*: Verzögerte freie Wiedergabe II, *T*: Trefferrate bei Wiedererkennen von Liste A, *FP*: Falsch Positive bei Wiedererkennen von Liste A, *d'*: Diskriminationswert bei Wiedererkennen Liste A, *DW*: nonparametrische Diskriminabilität.

TAP-Alertness: *AoM*: Median der Reaktionszeiten ohne Warnton, *AoSD*: Standardabweichung der Reaktionszeiten ohne Warnton, *AoAuslassungen*: Reaktionsauslassungen ohne Warnton, *AmM*: Median der Reaktionszeiten mit Warnton, *AmSD*: Standardabweichung der Reaktionszeiten mit Warnton, *AmAuslassungen*: Reaktionsauslassungen mit Warnton, *AmAntizipationen*: Anzahl Antizipationen.

4.3.1 Ein- und Ausschlusskriterien

Die Voraussetzungen für die Studienteilnahme waren ausreichende Seh- und Hörfähigkeiten sowie ausreichende Deutschkenntnisse, um das Instruktionsverständnis zu gewährleisten. Auf Grund der Altersnormierung der verwendeten Testverfahren wurde eine Altersspanne zwischen 18 und 60 Jahren als weiteres Einschlusskriterium definiert. In Bezug auf Geschlecht und Bildungsabschluss lagen keine Einschränkungen vor. Als Ausschlusskriterien wurden die diagnostischen Grenzen der PVT festgesetzt: Intelligenzminderung, dementielle Erkrankungen, amnestisches Syndrom, Schädelhirntraumata, Schlaganfall und Korsakoff-Syndrom sowie post-traumatische Amnesie. Darüber hinaus durften keine klinisch bedeutsamen neurologischen oder psychiatrischen Erkrankungen (wie schwere depressive Episode, Schizophrenie, Substanzmittelabusus oder -abhängigkeit) vorliegen (Schmand et al., 2005b). Ein weiteres Ausschlusskriterium stellte ein unzureichendes Instruktionsverständnis im Manipulationscheck dar.

4.3.2 Rekrutierung und Untersuchungsbedingungen

Der Erhebungszeitraum erstreckt sich von Januar 2014 bis November 2015. Die Rekrutierung der Probanden erfolgte größtenteils innerhalb des erweiterten Bekanntenkreises der Doktorandin sowie aus dem Personal der Klinik für Psychiatrie und Psychotherapie des Klinikums rechts der Isar. Die Studienteilnahme nahm zwischen 90 und 180 Minuten in Anspruch, die durchschnittliche Untersuchungsdauer betrug 118 Minuten ($SD = 19.76$). Es bestanden keine statistisch bedeutsamen Gruppenunterschiede in den Bearbeitungszeiten zwischen den Probanden der Kontrollgruppe KG ($M = 120.91$, $SD = 21.30$) und den experimentellen Simulanten ES ($M = 115.76$, $SD = 18.03$) vor ($t(64) = -1.06$, $p > .05$). Die Erhebungen fanden auf Wunsch der Probanden entweder beim Probanden zu Hause (47.0%), im Untersuchungszimmer der AG Klinische und Experimentelle Neuropsychologie im Klinikum rechts der Isar der Technischen Universität München (12.1%), bei der Doktorandin zu Hause (18.2%) oder im Elternhaus der Doktorandin (22.7%) statt. Die KG und ES unterschieden sich nicht in den Untersuchungsorten ($\chi^2(3, N = 66) = .432$, $p > .05$). Bei den Untersuchungen außerhalb der Klinik wurde darauf geachtet, dass eine störungsfreie Arbeitsatmosphäre (ruhiger Arbeitsplatz, keine Mobiltelefone) gewährleistet war. Die computergestützten Verfahren wurden mit einem Laptop erhoben. Vor jeder Untersuchung wurde kontrolliert, dass alle Aufgaben gut sichtbar waren. Trotz der diversen Erhebungsorte wurden die Untersuchungsbedingungen so konstant wie möglich gehalten.

4.3.3 Beschreibung der Stichprobe der Analogstudie

Insgesamt wurden 73 Probanden untersucht. Auf Grund des Überschreitens der Altersgrenze von 60 Jahren mussten drei Probanden aus den weiteren Analysen ausgeschlossen werden, ebenso ein Proband auf Grund eines Drogenmissbrauchs in der Vergangenheit. Darüber hinaus wurden von drei weiteren Probanden die Ergebnisse der computergestützten *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* wegen technischer Schwierigkeiten nicht gespeichert. Da der Datenverlust auf Grund technischer Probleme zufällig erfolgte und die Ergebnisanalyse nur vollständige Datensätze berücksichtigte, wurde auf eine weitere Auswertung dieser Probanden verzichtet. Daraus ergab sich eine Gesamtstichprobe von 66 Probanden, mit jeweils 33 Probanden in der Kontrollgruppe (KG) und experimentellen Simulanten (ES).

4.3.3.1 Soziodemographische und klinische Daten

Die soziographischen und klinischen Probandendaten unterteilt in Gesamtstichprobe Kontrollgruppe (KG) und experimentelle Simulanten (ES) sind in Tabelle 8 aufgeführt.

Die Stichprobe setzte sich aus 36 weiblichen (54.5%) und 30 männlichen (45.5%) Studienteilnehmern zusammen. Das mittlere Alter der Gesamtstichprobe lag bei $M = 35.7$ Jahren ($SD = 10.46$, Spannweite = 19 - 60). Es zeigten sich keine signifikanten Gruppenunterschiede bezüglich Alter, Geschlecht, Familienstand und Muttersprache. Auf Grund geringer Zellverteilungen wurden die Gruppen teilweise zusammengefasst. Die Muttersprache aller Probanden war Deutsch. Alle Probanden hatten einen Schulabschluss und eine Berufsausbildung absolviert oder befanden sich in Ausbildung. Die Mehrheit der Probanden hatte einen Hochschul- (54.4%) oder Fachhochschulabschluss (13.6%). Laut der Anamnese somatischer und psychischer Erkrankungen waren 4.5% der Probanden ($n = 3$) zum Zeitpunkt der Untersuchung auf Grund einer Anpassungsstörung und Angststörungen (Examensangst, Agoraphobie mit Panikstörung) in psychotherapeutischer Behandlung. 12.1% der Probanden litten unter somatischen Erkrankungen wie Schilddrüsenfunktionsstörungen (6.1%), Allergien (4.5%) und einem hypereosinophilen Syndrom (1.5%). Nach Kopfverletzungen, kognitiven oder zerebralen Beeinträchtigungen befragt, gaben 10.6% der Probanden an, im Verlauf ihres Lebens eine Gehirnerschütterung gehabt zu haben. Aktuell litt kein Proband unter kognitiven Beeinträchtigungen. Medikamentöse Nebenwirkungen auf die kognitive Leistungsfähigkeit konnten ebenfalls ausgeschlossen werden.

Tabelle 8

Deskriptive Statistiken soziodemographischer Daten der Gesamtstichprobe und getrennt nach randomisierter Gruppenzuweisung.

	total (N = 66)		KG (n = 33)		ES (n = 33)		p
Alter (Jahre)							
Mittelwert	35.70		36.94		34.45		.339 ^a
Standardabweichung	10.46		10.67		10.27		
Spannweite	19-60		25-60		19-58		
Geschlecht	n	%	n	%	n	%	
Männlich	30	45.5	16	48.5	14	42.4	.621 ^b
Weiblich	36	54.5	17	51.5	19	57.6	
Familienstand	n	%	n	%	n	%	
Ledig	21	31.8	11	33.3	10	30.3	.566 ^{b,c}
Partnerschaft, nicht zusammenlebend	7	10.6	4	12.1	3	9.1	
Partnerschaft, zusammenlebend	17	25.8	8	24.2	9	27.3	
Verheiratet	17	25.8	8	24.2	9	27.3	
Verheiratet, aber getrennt	0	0	0	0	0	0	
Verwitwet	1	1.5	0	0	1	3.0	
Geschieden	3	4.5	2	6.1	1	3.0	
Schulabschluss/Berufsausbildung (BA)	n	%	n	%	n	%	
Kein Abschluss	0	0	0	0	0	0	.314 ^{b,d}
Hauptschulabschluss ohne / in BA	0	0	0	0	0	0	
Hauptschulabschluss mit BA	5	7.6	1	3.0	4	12.1	
Mittlere Reife ohne / in BA	0	0	0	0	0	0	
Mittlere Reife mit BA	8	12.1	5	15.2	3	9.1	
Fachhochschulreife/Abitur ohne BA	0	0	0	0	0	0	
Fachhochschulreife/Abitur in BA	4	6.1	1	3.0	3	9.1	
Fachhochschulreife/Abitur mit BA	4	6.1	4	12.1	0	0	
Fachhochschulabschluss	9	13.6	3	9.1	6	18.2	
Hochschulabschluss	36	54.5	19	57.6	17	51.5	
Klinische Daten	n	%	n	%	n	%	
Psychische Störungen aktuell	3	4.5	0	0	3	9.1	.238 ^e
Somatische Erkrankungen aktuell	9	13.6	3	9.1	6	18.2	.475 ^e
Regelmäßige Medikation	15	22.7	6	18.2	9	27.3	.558 ^e
Kopfverletzungen lifetime	7	10.6	3	9.1	4	12.1	1.000 ^e

Anmerkungen.

^a t-Test für unabhängige Stichproben.

^b Chi-Quadrat-Test nach Pearson.

^c Berechnung des Chi-Quadrat-Tests anhand drei zusammengefasster Kategorien: („feste Partnerschaft nicht zusammenlebend“, „feste Partnerschaft zusammenlebend“) vs. („verwitwet“, „geschieden“, „verheiratet, aber getrennt“, „ledig“) vs. („verheiratet“).

^d Berechnung des Chi-Quadrat-Tests anhand drei zusammengefasster Kategorien: („kein Schulabschluss“, „Hauptschule mit/ ohne/ in BA“, „Mittlere Reife mit/ ohne/ in BA“) vs. („Fachhochschulreife/ Abitur mit/ ohne/ in BA“, „Fachhochschulabschluss“) vs. („Hochschulabschluss“).

^e exakter Test nach Fisher.

4.3.3.2 Weitere Einflussgrößen

Ergänzend wurden die subjektive Einschätzung der Anstrengungs- und Simulationsbereitschaft und weitere Einflussgrößen in neuropsychologischen Testsituationen (Müdigkeit, Langeweile, Interesse und Misstrauen) sowie spezifische Simulationsstrategien anhand visueller Analogskalen erfasst. Die Gruppen unterschieden sich nicht hinsichtlich Müdigkeit und Misstrauen, allerdings erlebten die experimentellen Simulanten (ES) während der Untersuchung mehr Langeweile und entwickelten weniger Interesse als die Kontrollgruppe (KG) (Tabelle 9).

Tabelle 9

Deskriptive Statistiken experimenteller Einflussvariablen.

	Total ($n = 66$)		KG ($n = 33$)		ES ($n = 33$)		t	df	p
	M	SD	M	SD	M	SD			
Müdigkeit	25.70	21.02	23.67	19.65	27.73	22.41	-.783	64	.437
Langeweile	21.95	24.09	13.30	18.62	30.61	26.03	-3.106	58	.003**
Interesse	80.89	21.37	90.36	11.30	71.61	24.88	3.944	45	.000***
Misstrauen	13.12	20.40	16.45	25.15	9.79	13.78	1.336	50	.188

Anmerkungen. ** $p < .01$, *** $p < .001$.

Die Motivation zur Studienteilnahme der Kontrollgruppe (KG) lag im Mittel bei 90.73 ($SD = 14.15$), die Anstrengungsbereitschaft während der Aufgabenbearbeitung betrug im Durchschnitt 94.64 ($SD = 7.62$). Analog zur Anstrengungsbereitschaft wurde die Simulationsbereitschaft der experimentellen Simulanten (ES) erfasst, die im Durchschnitt bei 84.82 ($SD = 16.04$) lag. Die Umsetzung der Instruktion, im Sinne einer Aufgabenbearbeitung entsprechend der Patientenrolle aus der Fallvignette, zeigte sich in einem Durchschnittswert von 86.09 ($SD = 14.18$). Des Weiteren wurden die Anstrengungsbereitschaft der KG und die Simulationsbereitschaft der ES in jedem einzelnen Verfahren anhand einer fünfstufigen Likertskala (0 „kaum vorhanden“ bis 4 „sehr stark vorhanden“) betrachtet (Tabelle 10).

Die Anstrengungsbereitschaft der KG in den einzelnen Verfahren zeigte, dass die KG bei der Bearbeitung des *Aggravations- und Simulationstests (AST)* die geringste und beim *California Verbal Learning Test (CVLT)* die größte Anstrengungsbereitschaft aufbrachten. Die Simulationsbereitschaft der ES zeigte beim *CVLT* ebenfalls die höchste Ausprägung. Die ES gaben zudem an, am wenigsten während der Bearbeitung des *Turms von London (TL-D)* simuliert zu haben. Die Einschätzung der angewandten Simulationsstrategien verdeutlichte, dass die ES sich am häufigsten nicht an alle Wörter erinnerten, die sie gewusst hätten ($M = 88.94$) und

mehr Fehler ($M = 88.06$) machten. Dagegen zeigten die Probanden seltener instruktionswidrige Verhaltensweisen ($M = 69.64$) und knapp-daneben Antworten ($M = 71.15$).

Tabelle 10

Deskriptive Statistiken der Anstrengungs- und Simulationsbereitschaft (fünfstufige Einschätzungen von 0 „kaum vorhanden“ bis 4 „sehr stark vorhanden“).

	AKGT	AST	WMT	RCFT	CVLT	TLD	ZVT	TAP: A	TAP: GA	TAP: GN
Anstrengungsbereitschaft KG ($n = 33$)										
<i>M</i>	2.97	2.30	2.55	2.88	3.42	3.12	2.82	2.79	3.18	2.94
<i>SD</i>	.81	1.05	.97	1.02	.66	1.02	1.16	.92	.81	.90
<i>Md</i>	3.00	2.00	3.00	3.00	4.00	3.00	3.00	3.00	3.00	3.00
<i>Spannweite</i>	1 - 4	0 - 4	2 - 4	0 - 4	2 - 4	0 - 4	0 - 4	1 - 4	1 - 4	1 - 4
Simulationsbereitschaft ES ($n = 33$)										
<i>Md</i>	2.76	2.67	2.94	2.52	3.00	2.12	2.30	2.85	2.61	2.76
<i>SD</i>	.97	1.02	.66	1.03	.63	1.22	1.19	1.28	1.03	1.17
<i>M</i>	3.00	3.00	3.00	3.00	3.00	2.00	2.00	3.00	3.00	3.00
<i>Spannweite</i>	1 - 4	0 - 4	1 - 4	0 - 4	2 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4

Ferner wurde der Einfluss der Reihenfolge der PVT auf die Testperformanz der ES untersucht. Die Testpermutation zeigte bei der Bearbeitung des *AKGT* einen statistisch bedeutsamen Unterschied ($F(5, 27) = 3.462, p < .015, \eta^2 = .391$). Die Post-hoc-Tests mit Bonferroni-Korrektur ergaben, dass die ES in der Testfolge 4 ($M = 80.40, SD = 5.32$) eine bessere Testperformanz als in Testfolge 6 ($M = 58.60, SD = 13.37$) präsentierten. Es zeigten sich keine Unterschiede in den sechs Testfolgen im *AST* ($F(5, 27) = 2.111, p < .095, \eta^2 = .281$), *WMT IR* ($F(5, 27) = 1.375, p < .265, \eta^2 = .203$), *WMT DR* ($F(5, 27) = 1.166, p < .351, \eta^2 = .178$) und folglich *WMT Cons* ($F(5, 27) = .261, p < .931, \eta^2 = .046$).

4.4 Statistische Datenanalysen

4.4.1 Datenaufbereitung

Im ersten Schritt der Datenaufbereitung wurden die Rohwerte der kognitiven Leistungstests anhand der jeweiligen Testhandbücher und Computerprogramme ausgewertet. Die Parameter wurden mit Normwerten (T -Werte, Prozentränge) verglichen und zur besseren Vergleichbarkeit einheitlich in T -Werte ($M = 50; SD = 10$) transformiert. Wenn in den Manualen keine T -Werte sondern Prozentränge angegeben waren, wurden diese mittels der

Transformationstabelle für Testnormen nach Lienert (1969), unter Berücksichtigung der Standardnormalverteilung, umgerechnet. Die computergestützte *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* ermöglichte eine automatische Auswertung und Umrechnung in *T*-Werte. Die Auswertung des *California Verbal Learning Test (CVLT)* wurde mit dem computergestützten Auswertungsprogramm des *Wiener Testsystems (WTS: Schuhfried, 2004)*, das nach Alter und Schulbildung (Hauptschule, Realschule, Gymnasium) normierte *T*-Werte ermittelt. Der Untertest *Kopie des Rey Complex Figure Test and Recognition Trial (RCFT)* wurde wegen besserer Differenzierungsgenauigkeit mit den Normwerten von Fastenau, Denburg und Hufford (1999) anstelle der Normen des Testhandbuches ausgewertet. Somit lagen für sämtliche kognitive Leistungsparameter alters- und evtl. auch geschlechts- und bildungskorrigierte *T*-Werte vor. Die Testleistungen in den PVT wurden als Rohwerte (*Amsterdamer Kurzzeitgedächtnistest AKGT*) bzw. Prozentwerte (*Aggravations- und Simulationstests AST, Word Memory Test WMT*) mit den entsprechenden Schwellenwerten verglichen. Auch die Angaben des postexperimentellen Fragebogens wurden als Roh- oder Prozentwerte angegeben.

Anschließend erfolgte die Dateneingabe in das Statistik-Programm *IBM SPSS Statistics* (Version 24.0 bis 26.0). Zunächst wurden die Vollständigkeit und Plausibilität der Daten geprüft. Die Verteilungsformen sämtlicher Variablen der beiden experimentellen Gruppen wurden jeweils anhand separater graphischer Darstellungen (Histogramme, Q-Q-Diagramme, Box-Plots) interpretiert. In den PVT waren die Verteilungsformen der experimentellen Simulanten (ES) mehrheitlich annähernd normalverteilt. Die Verteilungen der Kontrollgruppe (KG) waren meist linksschief und rechtssteil. In den Standardverfahren verliefen die Verteilungen der ES überwiegend rechtsschief und linkssteil, dagegen waren die Testwerte der KG meist normalverteilt oder linksschief und rechtssteil. Neben den graphischen Interpretationen wurden auch der Kolmogorov-Smirnov-Test mit Signifikanzkorrektur nach Lilliefors sowie der Saphiro-Wilk-Test zur Überprüfung der Normalverteilung herangezogen. Aufgrund seiner höheren Teststärke empfehlen mehrere Autoren wie Thode (2002) sowie Ghasemi und Zahediasl (2012) die Verwendung des Shapiro-Wilk-Tests, der den Zusammenhang zwischen der geschätzten Varianz einer Stichprobe und der erwarteten Varianz für normalverteilte Daten prüft. Die Verteilungen der beiden experimentellen Stichproben zeigten diverse Ausreißerwerte und Extremwerte. Neben (milden) Ausreißern zwischen 1.5 und 3 Interquartilen lagen auch extreme Ausreißer (Extremwerte) von mindestens 3 Interquartilen vor. Nach Field (2013) sind insbesondere extreme Ausreißer zu

kontrollieren, während für milde Ausreißer eine gewisse Toleranz in den Statistiken angenommen werden kann. Jeder Wert wurde bzgl. fehlerhafter Auswertung und Eingabe kontrolliert. Anschließend wurde eine Korrektur der extremen Ausreißerwerte im Sinne einer Winsorisierung (Tukey, 1962) in Betracht gezogen. Dabei wurden die Extremwerte jeweils durch den nächstgrößeren bzw. kleineren Wert der Verteilung ersetzt (Sachs & Hedderich, 2009). Es zeigte sich, dass die Gruppenunterschiede nicht auf Extremwerten basierten oder entsprechend verzerrt wurden. Da die Extremwerte in inhaltlich plausiblen und nachvollziehbaren Wertebereichen rangierten, wurden diese der insgesamt 10 Variablen beibehalten. Bei der folgenden Methodenauswahl wurden robuste Verfahren oder ein vergleichendes Vorgehen gewählt, um möglichen Verzerrungen entgegenwirken zu können.

4.4.2 Methodenauswahl und Variablenselektion

In dieser Arbeit wurden als Maße der zentralen Tendenz arithmetische Mittel und als Dispersionsstatistiken Standardabweichungen gewählt, da sich keine Abweichungen zu non-parametrischen Analysen mit Medianwerten zeigten. Das Signifikanzniveau wurde für alle Analysen auf $\alpha = .05$ (zweiseitige Testung) festgesetzt.

Zunächst wurden die Mittelwerte sämtlicher Leistungsparameter, einschließlich 95%-Konfidenzintervalle, deskriptiv mittels Excel graphisch dargestellt. Dies ermöglichte eine visuelle Veranschaulichung der Verteilung der Stichproben in Kapitel 7.1 (*Abbildung 3*, *Abbildung 4*). Zur Analyse bedeutsamer Gruppenunterschiede wurde zunächst die Verwendung multivariater Varianzanalysen (MANOVAs) und multipler *t*-Tests gegeneinander abgewogen. Als Vorteile einer MANOVA gelten das Vermeiden einer α -Fehler-Kumulierung und die Berücksichtigung von Innergruppeneffekten der Zusammenhänge zwischen den abhängigen Variablen (Field, 2013). Somit wurden zunächst die Voraussetzungen geprüft. Aufgrund der randomisierten Stichprobenzuteilung war die Unabhängigkeit der Beobachtungen gegeben. Zur Überprüfung der multivariaten Normalverteilung wurden die univariaten Verteilungen herangezogen (Bortz & Schuster, 2010). Die Homogenität der Varianz-Kovarianz-Matrizen wurde mittels Box-M-Tests kontrolliert. Die Multikollinearität der abhängigen Variablen, im Sinne hoher Interkorrelationen der AV, können in Fehlinterpretationen der Ergebnisse resultieren. Die Interkorrelationen sollten dabei nicht größer als $r(Y_1, Y_2) > 0.70$ sein. Zudem wurden die Kennwerte VIF (*variance inflation factor*) und Toleranz ermittelt, wobei der VIF den Wert von 5.00 nicht überschreiten und die Toleranz den Wert von 0.25 nicht unterschreiten sollte (Urban & Mayerl, 2011). Die VIF-Werte rangierten für die PVT zwischen 14.84 (*WMT Cons*

und *WMT DR*) und 3.42 (*AST* und *WMT IR*). Auch die VIF-Werte der neuropsychologischen Leistungsparameter überschritten mehrheitlich den Grenzwert von 5.00, teilweise um ein Vielfaches. Auf Grund dieser Verletzungen und der Sensitivität der Teststatistik gegenüber Ausreißerwerten wurden multiple *t*-Tests für unabhängige Stichproben berechnet. Bei Varianzheterogenität wurden die entsprechenden SPSS-korrigierten Werte verwendet (Rasch et al., 2010). Die Justierung des α -Fehlers der multiplen *t*-Tests erfolgte anhand einer sequentiellen Bonferroni-Korrektur nach Holm. Dabei wurden die einzelnen Kennwerte der Größe nach absteigend hinsichtlich des jeweils modifizierten Signifikanzniveaus getestet (Bortz & Schuster, 2010). Zur Veranschaulichung signifikanter Mittelwertdifferenzen wurden zudem die Effektstärken der Gruppenunterschiede anhand gepolter Standardabweichungen nach Cohens *d* (Cohen, 1988) ermittelt. Auf Grund der fehlenden Normalverteilung wurden die Ergebnisse mit den Ergebnissen non-parametrischer Mann-Whitney-*U*-Tests verglichen. Da sich keine bedeutsamen Abweichungen zeigten, werden in der folgenden Ergebnisdarstellung nur die parametrischen Auswertungen aufgeführt.

Darüber hinaus wurde der Einfluss der subjektiven Schwierigkeit auf die Testperformanz untersucht. Da die Einschätzung der subjektiven Schwierigkeit auf Ordinalskalenniveau vorlag, wurde auch hier ein vergleichendes Vorgehen von parametrischen und nonparametrischen Analysen gewählt. Neben den Unterschieden zwischen den Gruppen hinsichtlich der subjektiven Schwierigkeiten (*t*-Tests für unabhängige Stichproben), wurden auch die Unterschiede zwischen den einzelnen PVT separat für beide Gruppen (*t*-Tests für verbundene Stichproben) ermittelt. Darüber hinaus wurde der Zusammenhang zwischen Schwierigkeit und Testperformanz untersucht. Neben Korrelationskoeffizienten nach Pearson für intervallskalierte Merkmale, wurden auch Rangkorrelationen nach Spearman berechnet. Auf Grund vergleichbarer Ergebnisse und der Äquidistanzen zwischen den 5 Stufen der Selbstbeurteilung, wurden die Koeffizienten nach Pearson angegeben.

Zur Spezifikation der Klassifikationsgüte der eingesetzten Verfahren wurde im nächsten Schritt zwischen der Verwendung binär-logistischer Regressionsanalysen und linearer Diskriminanzanalysen abgewogen. Die Teststatistik einer linearen Diskriminanzanalyse entspricht einer inversen MANOVA, deren Voraussetzungen sich überschneiden. Lineare Diskriminanzanalysen ermöglichen Schätzungen mit höherer Güte und Effizienz. Allerdings birgt eine zu hohe Heterogenität der Varianz-Kovarianz-Matrizen, insbesondere bei klassifikatorischen Diskriminanzanalysen, das Risiko falscher Zuordnungen in die Gruppe mit der größeren Varianz (Rudolf & Müller, 2012). In dieser Arbeit wurden binär-logistische

Regressionsanalysen auf Grund ihrer größeren Robustheit verwendet. Zu den Voraussetzungen zählt die Binomialverteilung der Kriteriumsvariablen, die automatisch durch dichotome Ausprägungsmöglichkeiten (0 oder 1) erfüllt wird. Zudem wird die statistische Unabhängigkeit dieser dichotomen Kriteriumsausprägung vorausgesetzt, die durch eine randomisierte Gruppenzuteilung gewährleistet wird. Des Weiteren wird das Fehlen leerer Zellen der verschiedenen Prädiktor-Kriterium-Kombinationen angenommen (Rudolf & Müller, 2012). In binär-logistischen Regressionsanalysen wird die Wahrscheinlichkeit eines dichotomen Kriteriums aus nominal- oder intervallskalierten Prädiktorvariablen berechnet. Während in linearen Regressionsgleichungen die optimalen Schätzungen der Regressionskoeffizienten anhand der Methode der kleinsten Quadrate berechnet werden, um den Einfluss der erklärenden Variablen auf die abhängige Variable direkt zu ermitteln, basieren logistische Regressionen auf *Maximum-Likelihood-Schätzungen* (M-L-Schätzungen). In einem iterativen Vorgehen werden die optimalen Koeffizienten ermittelt, um den Einfluss der erklärenden Variablen auf die Wahrscheinlichkeit des Auftretens des Wertes eins der abhängigen Variablen zu schätzen. Auf Grund der logistischen Verteilungsfunktion rangieren die Prognosen des Regressionsmodells zwischen null und eins und können als Wahrscheinlichkeiten interpretiert werden. Die iterative Schätzung des Vorsagemodells wird abgebrochen, sobald der Zuwachs des vorherigen Kriteriums unterschritten wird. In der vorliegenden Arbeit wurde die Modellpassung mit dem Hosmer-Lemeshow Test überprüft. Der Hosmer-Lemeshow Test unterteilt eine Stichprobe in maximal 10 Gruppen und ermittelt die Differenzen zwischen den beobachteten und erwarteten Werten. Kleinere Differenzen gehen mit einer besseren Modellpassung einher.

Die Aussagekraft des binär-logistischen Regressionsmodells hängt von einer abgeschlossenen M-L-Schätzung ab und wird durch eine (quasi-)vollständige Trennung (Separation) der Variablen, im Sinne einer eindeutigen Aufteilung der Variablen, reduziert. Als *vollständige Separation* wird die perfekte Gruppeneinteilung durch einen Prädiktor bezeichnet. Eine *quasi-vollständige Separation* liegt vor, wenn mindestens ein Proband auf den Grenzbereich zwischen den beiden Gruppen fällt (Albert & Anderson, 1984). In Kreuztabellen wird eine Separation durch die fehlende Besetzung aller Prädiktorwertekombinationen ersichtlich, in logistischen Regressionsmodellen durch Parameterschätzungen, die mit großen Standardfehlern und kleinen z-Werten im Wald-Test einhergehen. Heinze und Schemper (2002) sowie Wollschläger (2014) empfehlen eine Korrektur nach Firth, im Sinne einer *penalisierten Likelihood-Methode*. Durch die Integration eines Strafterms können die

resultierenden unendlichen Parameterschätzer in der *M-L*-Schätzung korrigiert werden. Anhand des Erweiterungsmoduls „STATS_ FIRTHLOG“ Version 1.1.0 für IBM SPSS Statistics 24 wurden die korrigierten β -Gewichte ermittelt und zur Berechnung korrigierter logistischer Regressionsanalysen nach Firth herangezogen. Die Berechnung der Wahrscheinlichkeit der Gruppenzugehörigkeit in Excel erfolgte mit den korrigierten β -Gewichten in der logistischen Regressionsgleichung. Als Klassifikations-schwellenwert wurde entsprechend der üblichen Vorgehensweise $p(y) = .50$ festgelegt.

$$p(y = 1) = \frac{1}{1 + e^{-(b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni})}} = \frac{e^{(b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni})}}{1 + e^{(b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni})}}$$

Ergänzend wurden als Maß für die zufallskorrigierte Übereinstimmung der Methoden Kappa-Koeffizienten κ nach Cohen berechnet. Gemäß Altman (1990) gilt die Übereinstimmung bei Kappa-Koeffizienten von < 0.2 als sehr schlecht, zwischen 0.21 und 0.40 als schlecht, zwischen 0.41 und 0.6 als mäßig, zwischen 0.61 und 0.8 als gut und zwischen 0.81 und 1.0 als sehr gut.

Im nächsten Schritt wurden die Klassifikationsstatistiken anhand *Receiver Operating Characteristics* (ROC) Analysen bestimmt. Das verteilungsfreie Verfahren entstammt der Signalentdeckungstheorie der Psychophysik. Es eignet sich zur Gruppeneinteilung eines dichotomen Kriteriums unabhängig der Gruppengrößen. In einer ROC-Kurve werden die Sensitivitätswerte auf der Y-Achse und die Werte 1-Spezifität auf der X-Achse abgebildet. Eine Gruppentrennung auf Zufallsniveau würde der Hauptdiagonale folgen, so dass sich mit steigender Trennschärfe der stichprobenabhängige ROC-Kurvenverlauf den oberen Seiten des umschließenden Quadrates annähert (Goldhammer & Hartig, 2012). Der optimale Schwellenwert lässt sich anhand des Youden-Index J (Sensitivität + Spezifität – 1) im Wertebereich zwischen 0 und 1 berechnen. Die Fläche unterhalb der ROC-Kurve, die *Area under the Curve* (AUC) entspricht exakt der Wahrscheinlichkeit einer korrekten Klassifikation des Verfahrens (Wertebereich 0.5 bis 1). Hosmer und Lemeshow (2000) empfehlen für die Interpretation der AUC-Werte folgende Richtwerte: 0.5 entspricht einer Differenzierung nach Zufall, $< .60$ wird als wertlos, $.61$ bis $.70$ als schlecht, $.71$ bis $.80$ als akzeptabel, $.81$ bis $.90$ als ausgezeichnete und $\geq .90$ als herausragend eingeteilt. Äquivalent zur Wilcoxon-Rangsummen-Teststatistik kann anhand non-parametrischer Mann-Whitney *U*-Tests überprüft werden, ob sich die AUC von der Hauptdiagonalen (Zufallswahrscheinlichkeit) unterscheidet. Im Gegensatz zu bisherigen Forschungsarbeiten wurden in dieser Studie *T*-Werte statt Rohwerte

zur Ermittlung von Schwellenwerten bei den eingebetteten Testparametern verwendet. Da die Beurteilung der Testperformanz in der neuropsychologischen Praxis anhand von Normwerten erfolgt, wurden auch hier die Gruppenvergleiche und damit konsistent die Schwellenwerte mit normierten T -Werten berechnet. Die Auswahl optimaler statistischer Schwellenwerte erfolgte unter Berücksichtigung des Youden-Indexes J und der BV-Konventionen (Spezifität $\geq 90\%$, Sensitivität $\geq 50\%$). Insofern sich bei der Berechnung nach Youden's J gleiche Schwellenwerte ergaben, erfolgte die Auswahl nach der größtmöglichen Spezifität. Neben Schwellenwerten wurden auch die positiven Vorhersagevaliditäten (PVV) und negativen Vorhersagevaliditäten (NVV) unter Berücksichtigung der Prävalenzraten bestimmt. Ferner wurden auch die *Odds Ratios* (OR), als Quotenverhältnis der Treffer und korrekten Ablehnungen zu den Verpassern und falschen Alarmen und die *Klassifikationsgenauigkeit* als *Trefferrate* (*Hit-Rate*, HR) ermittelt.

Abschließend wurde zum Vergleich diversen Vorhersagemodelle neben den Kappa- und AUC-Werten auch zwei weitere Kriterien zur Beurteilung der Modellgüte berechnet. Um einen Kompromiss zwischen hohem Erklärungsgehalt und Parameterzahl zu bestimmen, wurde für jedes Vorhersagemodell der *Akaike Information Criterion* AIC-Wert ermittelt. Dabei sollen Modelle mit kleineren AIC-Werten ausgewählt werden. Ferner wurden die *Bayesian Information Criterion* BIC-Werte bestimmt, welcher im Gegensatz zum AIC den Stichprobenumfang berücksichtigt.

5 Methodik: zweiter Teil der Studie

Im zweiten Teil der Studie erfolgt in einem partiellen Stichprobenvergleich eine Gegenüberstellung der experimentellen Simulanten (ES) der Analogstudie und Patienten mit negativen Antwortverzerrungen (PNA). In diesem Abschnitt werden neben den Untersuchungsbedingungen auch die Stichprobe und die statistischen Analysen beschrieben.

5.1 Untersuchungsbedingungen

Die Patienten mit negativen Antwortverzerrungen (PNA) wurden im Rahmen der konsiliarischen Tätigkeit der Arbeitsgruppe Klinische und Experimentelle Neuropsychologie der Klinik für Psychiatrie und Psychotherapie des Klinikums rechts der Isar zwischen Juli 2012 und Juni 2016 untersucht. Sämtliche Untersuchungen wurden von der Doktorandin in den Räumlichkeiten der Arbeitsgruppe durchgeführt. Die neuropsychologische Untersuchung der Patienten beinhaltete jeweils eine Erfassung des kognitiven Leistungsniveaus und eine Einschätzung der Beschwerdvalidität der ermittelten kognitiven Defizite. Die Überprüfung der Beschwerdvalidität erfolgte anlässlich eines entsprechenden Verdachts des behandelnden Arztes oder eines Begutachtungsauftrages (Versicherungsträger, staatlicher Arbeitgeber, Gericht). Während die experimentellen Simulanten (ES) gezielt für die vorliegende Arbeit rekrutiert und untersucht wurden, erfolgten die Erhebungen der PNA im Kontext der regulären Behandlung und des klinischen Versorgungsauftrags der Arbeitsgruppe. Die Daten resultierten somit nicht aus einer experimentellen Manipulation und wurden in retrospektiven Analysen in diese Arbeit aufgenommen. Demzufolge lagen keine entsprechenden Einverständniserklärungen vor. Laut des Rechtsgutachtens zum Datenschutz in der medizinischen Forschung (Dierks, Roßnagel, Hornung & Jandt, 2009) können bei strikter Trennung der Identitätsdaten von den Patientendaten entsprechende Daten verwendet werden. Dabei muss der Ausschluss eines Personenbezugs, beispielsweise im Sinne einer Anonymisierung, gewährleistet sein. „Handelt es sich um nicht personenbezogene Behandlungsdaten, fallen sie nicht in den Anwendungsbereich der Datenschutzgesetze und können ohne weitere Voraussetzungen für medizinische Forschungszwecke eingesetzt werden.“ (Dierks et al., 2009, S. 27).

Die neuropsychologischen Untersuchungen fanden jeweils an einem Tag statt und dauerten, ohne Anamnesegespräch zwischen 120 und 300 Minuten, wobei die durchschnittliche Dauer 219 Minuten betrug. Insgesamt lagen 28 Datensätze vor. In die Analysen wurden nur

Datensätze mit allen drei PVT, den drei Untertests der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* und des *California Verbal Learning Tests (CVLT)* eingeschlossen. Die Testverfahren wurden bereits in Abschnitt 4.2 beschrieben. Da die Verfahren Bestandteil einer vollständigen neuropsychologischen Untersuchung waren, wurde auf eine Kombination zusätzlicher Leistungstests entsprechend der Analogstudie verzichtet. Auf Grund fehlender Werte konnten sechs Datensätze (21.4%) nicht berücksichtigt werden. Zwei weitere Datensätze (7.1%) wurden wegen Überschreiten der Altersgrenze von 60 Jahren ausgeschlossen. Zudem musste mindestens ein auffälliger Wert im *Amsterdamer Kurzzeitgedächtnistest (AKGT)* oder im *Word Memory Test (WMT)* vorliegen. Um Fehlklassifikationen durch eine Konfundierung tatsächlicher kognitiver Dysfunktionen zu vermeiden, wurden fünf Patienten (17.9%) mit organischen oder schweren psychischen Erkrankungen und negative Antwortverzerrungen ausgeschlossen. Darüber hinaus erzielten fünf weitere Patienten (17.9%) weder in einem der PVT noch in den Standardverfahren auffällige Leistungen und wurden ebenfalls nicht in die Analysen einbezogen. In die Studie wurden nur Patienten eingeschlossen, deren Anstrengungsbereitschaft durch die Arbeitsgruppe als fragwürdig beurteilt wurde.

Schließlich lag die finale Stichprobengröße der Patienten mit negativen Antwortverzerrungen (PNA) bei $n = 10$ Patienten, was 35.7% der Gesamtstichprobe entsprach. Die Stichprobe setzte sich aus sieben ambulanten Gutachtenpatienten (70%) und drei stationären Patienten (30%) zusammen, deren Anstrengungsbereitschaft als fragwürdig beurteilt wurde. Von den PNA erzielten 70% sowohl im *AKGT* als auch im *WMT* auffällige Werte. 20% der Patienten präsentierten lediglich im *WMT* auffällige Werte und 10% nur im *AKGT*. Ein Patient erzielte im *AKGT* und *WMT* Ergebnisse, die signifikant unterhalb der Ratewahrscheinlichkeit lagen, so dass in diesem Fall von einer bewussten Antwortmanipulation ausgegangen werden konnte. Drei weitere Patienten demonstrierten im *WMT* eine Testperformanz innerhalb des Zufallsbereichs ($34.5\% \leq WMT \leq 65.5\%$). Unter Berücksichtigung der diagnostischen Sicherheit kann die Stichprobe in sichere Simulanten ($n = 1$), wahrscheinliche Simulanten ($n = 6$) und mögliche Simulanten ($n = 3$) differenziert werden, die für die inferenzstatistischen Analysen als PNA zusammengefasst wurden.

5.2 Beschreibung der Stichprobe PNA

Die Stichprobe der PNA setzte sich aus 6 männlichen (60%) und 4 weiblichen (40%) Patienten zusammen. Das mittlere Alter lag bei $M = 42.2$ Jahren ($SD = 9.91$, Spannweite = 22 – 52).

Neun von zehn Patienten (90%) waren deutsche Muttersprachler. Mit Ausnahme eines Patienten hatten alle Patienten eine Berufsausbildung absolviert. Jeweils vier Patienten (40%) hatten entweder einen Hauptschulabschluss oder mittlere Reife. Ein Patient hatte einen Fachhochschul- bzw. Hochschulabschluss. Neun Patienten (90%) litten an einer psychischen Störung, sieben (70%) zeigten komorbide Störungsbilder. Die häufigsten Diagnosegruppen waren mit jeweils fünf Patienten affektive Störungen (F30) und somatoforme Störungsbilder (F45). Zwei Patienten (20%) litten an Angst- und Belastungsreaktionen (F40-43), ein Patient (10%) an einer Persönlichkeitsstörung (F60) sowie Wesensveränderungen auf Grund organischer Ursachen (F00). An somatischen oder neurologischen Erkrankungen waren jeweils zwei Patienten (20%) erkrankt. Die neurologischen Störungen waren Migräne und Paresen. Die Patienten der klinischen Stichprobe litten somatisch zudem an Störungen des Ganges, Extremitätenataxie, Z. n. Nackenwirbelsäulen-Distorsion und Krankheiten des Ohres. Sieben Patienten (70%) nahmen regelmäßig verschreibungspflichtige Medikamente ein, sechs Patienten (60%) Psychopharmaka.

Um die Vergleichbarkeit der Patienten mit negativen Antwortverzerrungen (PNA) und den experimentellen Simulanten (ES) zu ermitteln, werden in diesem Abschnitt die soziographischen und klinischen Daten der beiden Stichproben in Tabelle 11 dargestellt. Zwischen den beiden Stichproben wurden keine statistisch bedeutsamen Unterschiede hinsichtlich Geschlecht, Familienstand und dem Vorliegen somatischer Erkrankungen ermittelt. Dagegen zeigten sich signifikante Unterschiede in Alter, Bildungsabschluss und dem Vorliegen neurologischer und psychischer Erkrankungen, einschließlich regelmäßiger Medikation. Während die PNA im Durchschnitt 42.20 Jahre alt waren, lag das mittlere Alter der ES mit 34.45 Jahren unterhalb der PNA. Die Unterschiede zwischen den Gruppen im Ausbildungsniveau zeigten sich insbesondere in der Häufigkeit höherer Bildungsabschlüsse. Ein Patient der PNA (10.0%) hatte einen höheren Bildungsabschluss (Hochschulabschluss), dagegen lag bei 51.5% der ES ein Hochschulabschluss und bei 18.2% ein Fachhochabschluss vor. Auf Grund der Alters- und Bildungsnormierung der kognitiven Leistungstests, kann der Einfluss der Bildungs- und Altersdifferenzen als kontrolliert betrachtet werden. Ferner litten die PNA erwartungsgemäß signifikant häufiger an psychischen Erkrankungen und nahmen regelmäßig Medikamente, einschließlich Psychopharmaka.

Tabelle 11

Deskriptive Statistiken soziodemographischer Daten der PNA und ES, deren Gesamtstichprobe und Ergebnisse des statistischen Vergleichs der Teilstichproben.

	Total (N = 43)		PNA (n = 10)		ES (n = 33)		p
Alter (Jahre)							
Mittelwert	36.26		42.20		34.45		.041 ^a
Standardabweichung	10.60		9.91		10.27		
Spannweite (Range)	19 - 58		22 - 52		19 - 58		
Geschlecht							
	N	%	n	%	N	%	
Männlich	20	46.5	6	60.0	14	42.4	.473 ^b
Weiblich	23	53.5	4	40.0	19	57.6	
Muttersprache							
	N	%	n	%	N	%	
Deutsch	42	97.7	9	90.0	33	100	.233 ^b
Andere	1	2.3	1	10.0	0	0	
Familienstand							
	N	%	N	%	N	%	
Ledig	12	27.9	2	20.0	10	30.3	.149 ^{cd}
Feste Partnerschaft nicht	3	7.0	0	0	3	9.1	
Feste Partnerschaft zusammenlebend	10	23.3	1	10.0	9	27.3	
Verheiratet	12	27.9	3	30.0	9	27.3	
Verheiratet, aber getrennt	1	2.3	1	10.0	0	0	
Verwitwet	2	4.7	1	10.0	1	3.0	
Geschieden	3	7.0	2	20.0	1	3.0	
Schulabschluss/Berufsausbildung							
	N	%	n	%	N	%	
Kein Abschluss	0	0	0	0	0	0	.000 ^{c e}
Hauptschulabschluss ohne BA	1	2.3	1	10.0	0	0	
Hauptschulabschluss mit BA	8	18.6	4	40.0	4	12.1	
Mittlere Reife mit BA	7	16.3	4	40.0	3	9.1	
Fachhochschulreife/Abitur in BA	3	7.0	0	0	3	9.1	
Fachhochschulabschluss	6	14.0	0	0	6	18.2	
Hochschulabschluss	18	41.9	1	10.0	17	51.5	
Klinische Daten							
	N	%	N	%	N	%	
Psychische Störungen	12	27.9	9	90.0	3	9.1	.000 ^b
Neurologische Erkrankungen	2	4.7	2	20.0	0	0	.050 ^b
Komorbidität	9	20.9	7	70.0	2	6.1	.000 ^b
Somatische Erkrankungen	8	18.6	2	20.0	6	18.2	1.000 ^b
Regelmäßige Medikation	16	37.2	7	70.0	9	27.3	.024 ^b
Psychopharmaka	6	14.0	6	60.0	0	0	.000 ^b

Anmerkungen.

^a t-Test für unabhängige Stichproben.

^b Exakte Signifikanz nach Fisher.

^c Berechnung des Chi-Quadrat-Tests durch Bildung von zwei Kategorie Partnerschaft ja und Partnerschaft nein.

^d Chi-Quadrat nach Pearson.

^e Berechnung des Chi-Quadrat-Tests durch Bildung von drei Kategorien einfaches, mittleres und höheres Ausbildungsniveau.

5.3 Statistische Datenanalysen

Die Datenaufbereitung erfolgte entsprechend des Vorgehens der experimentellen Analogstudie und die Daten wurden in den bestehenden Datensatz in SPSS aufgenommen. Daraufhin erfolgte zunächst eine Überprüfung der Daten auf Vollständigkeit und Plausibilität. Box-Plot-Analysen verdeutlichten, dass keine Extremwerte von mindestens 3 Interquartilen vorlagen. Zur inferenzstatistischen Absicherung wurde die Normalverteilungsannahme geprüft. Analog zu den Verteilungsformen der experimentellen Simulanten (ES) waren die Ergebnisse der Patienten mit negativen Antwortverzerrungen (PNA) in den PVT weitgehend normalverteilt. Dagegen verliefen die Verteilungen in den Standardverfahren überwiegend rechtsschief und linkssteil. In Kapitel 8.1 erfolgte zunächst eine graphische Gegenüberstellung der Mittelwerte der ES und PNA (*Abbildung 7, Abbildung 8*). Anschließend wurden *t*-Tests mit Bonferroni-Korrektur nach Holm zur Überprüfung der Mittelwertdifferenzen durchgeführt, die keine Abweichungen zu non-parametrischen Auswertungen zeigten. Zur Validierung der Vorhersagemodelle und Schwellenwerte der Analogstudie wurden die PNA mit den ermittelten Kennwerten klassifiziert.

6 Methodik: dritter Teil der Studie

6.1 Studiendesign und Ablauf

In der vorliegenden Arbeit wurden auch neurologisch-psychiatrische Patienten mit kognitiven Beeinträchtigungen (PKB) untersucht. Die Patienten nahmen während einer klinischen Routineuntersuchung der AG Neuropsychologie im Klinikum rechts der Isar oder im Neurozentrum Starnberg an der Studie teil. Die PKB wurden im Vorfeld der neuropsychologischen Untersuchung schriftlich und mündlich über die Teilnahme an der Studie aufgeklärt und unterzeichneten eine schriftliche Einverständniserklärung. Nach der klinischen Anamnese und Bearbeitung der Testbatterie *Consortium to Establish a Registry for Alzheimer's Disease* (CERAD: Morris et al., 1989) der Routineuntersuchung erfolgte die zusätzliche Bearbeitung der *Kurzform der Geriatrischen Depressionsskala* (GDS-15, Sheikh & Yesavage, 1986). Darüber hinaus bearbeiteten die Patienten den *Aggravations- und Simulationstests* (AST) und die erste Testhälfte des *Amsterdamer Kurzzeitgedächtnistest* (AKGT). Das Studiendesign beinhaltet keine zusätzlichen experimentellen Variationen. Als Einschlusskriterien wurden in Anlehnung an Merten, Bossink, et al. (2007) ein Demenzverdacht, keine Anzeichen mangelnder Kooperationsbereitschaft, ein Minimum von acht Jahren formaler Schulausbildung und keine dokumentierte geistige Behinderung definiert. Weitere Kriterien stellten ein Mindestalter von 65 Jahren und ausreichende Deutschkenntnisse dar.

6.2 Untersuchungsverfahren

Die beiden PVT *Amsterdamer Kurzzeitgedächtnistest* (AKGT) und *Aggravations- und Simulationstests* (AST) wurden bereits in Abschnitt 4.2.3 beschrieben, so dass im Folgenden die beiden Verfahren aufgeführt werden, die nur die PKB bearbeiteten.

Consortium to Establish a Registry for Alzheimer's Disease (CERAD: Morris et al., 1989)

Morris und Kollegen des amerikanischen *National Institute of Aging* entwickelten 1989 zur standardisierten Diagnostik kognitiver Beeinträchtigungen der Alzheimer-Demenz die Testbatterie der CERAD. Das zeitökonomische Verfahren ermöglicht neben einer Differenzierung der kognitiven Leistungsfähigkeit und des Schweregrads kognitiver Funktionsstörungen auch eine Verlaufskontrolle therapeutischer Interventionen (Rabitsch, 2014). Im Rahmen der Autorisierung der *Memory Clinic* der Geriatrischen Universitätsklinik

Basel entstand die deutschsprachigen Version *CERAD-NTB* (Thalman et al., 2000), die durch drei weitere Subtests zur *CERADplus-NTB* ergänzt wurde. Neben der Erfassung der Gedächtnis- und Sprachfunktionen, Orientierung und Praxie umfasst die Version weitere Exekutivfunktionen. Dadurch wurde das diagnostische Spektrum für andere Demenzformen, wie frontotemporale Demenzen, ergänzt (Jahn & Werheid, 2015). Die Untersuchungsmaterialien und eine computergestützte Auswertung per Excel- oder Onlinedatei sind kostenfrei unter <https://www.memoryclinic.ch/de/> erhältlich. In dieser Arbeit wurde die erweiterte deutschsprachige Version *CERAD-Plus test battery, revised edition* (Memory Clinic University Hospital Basel, 2005) verwendet.

Durchführung: Die *CERADplus-NTB* besteht aus 11 Subtests, aus denen 16 Kennwerte ermittelt werden können. Die erste Aufgabe 1) *Verbale Flüssigkeit: Kategorie 'Tiere' (VF)* erfasst die Sprachproduktion, semantische Abrufleistungen und exekutive Funktionen (Rabitsch, 2014). Die Testperson wird dabei gebeten, so viele Tiere wie möglich innerhalb einer Minute aufzuzählen. 2) Mit den 15 Objekten des *Boston Naming Tests (BNT)* werden die visuelle Wahrnehmung, das Wortbenennen und die Wortfindung untersucht. Der Proband wird aufgefordert den Namen von gezeichneten Objekten zu benennen. 3) Der *Mini-Mental State Examination (MMSE)*: Folstein et al., 2000) stellt ein kurzes Screeningverfahren diverser Funktionsbereiche wie Orientierung, Sprache, Gedächtnis Aufmerksamkeit und Praxie dar. 4) Die verbale Lernfähigkeit wird mit der *Wortliste Lernen (WLG)* erfasst. Eine Wortliste mit 10 Begriffen soll nach drei *Lerndurchgängen (DG1, DG2, DG3)* jeweils aus dem Gedächtnis wiedergegeben werden, um den Lernzuwachs zu erfassen. Die Nennung zusätzlicher Begriffe, die nicht in der Wortliste enthalten waren, werden als Intrusionen bezeichnet und im Kennwert *Wortliste Intrusionen (WLI)* summiert. 5) Die visuell-räumlichen Fähigkeiten werden durch Abzeichnen von vier geometrischen Figuren verschiedener Schwierigkeit im Untertest *Figuren Abzeichnen (FAZ)* (Rosen, Mohs & Davis, 1984) erfasst. 6) Anschließend wird das verbale episodische Gedächtnis, die mittelfristige verbale Merkfähigkeit, anhand der zeitlich verzögerten Wiedergabe der zuvor dargebotenen 10 Wörter mit den Untertests *Wortliste Abrufen (WLA)* und *Wortliste Behalten (WLB)* ermittelt. 7) Die *Wortliste Wiedererkennen (WLW)* überprüft die Rekognitionsleistung der 10 gelernten Begriffe, indem diese Begriffe aus 10 weiteren, neuen Begriffen wiedererkannt werden sollen. Aus dem Verhältnis zwischen der Differenz richtig wieder erkannter und richtig als neu erkannter Wörter sowie der Gesamtzahl an 20 Wörtern lässt sich ein Kennwert der *Diskriminabilität (D)* berechnen. 8) Die figurale Gedächtnisleistung wird durch die Zeichnung der geometrischen Figuren aus dem Gedächtnis

mit den Kennwerten *Figuren Abrufen (FA)* und *Figuren Behalten (FB)* ermittelt. Die *CERADplus-NTB* beinhaltet zudem die Untertests 9) *Phonematische Flüssigkeit: S-Wörter* sowie den 10) *Trail-Making-Test Teil A (TMT-A)* und 11) *Trail-Making-Test Teil B (TMT-B)*. Im *TMT-A* werden die Ziffern von 1 bis 25 in aufsteigender Reihenfolge so schnell wie möglich mit einer durchgehenden Linie verbunden. Dadurch werden die Informationsverarbeitungs-geschwindigkeit und gerichtete Aufmerksamkeit geprüft. Die Exekutivfunktionen, insbesondere die geteilte Aufmerksamkeit und kognitive Flexibilität, misst der *TMT-B*, indem alternierend Zahlen und Buchstaben in aufsteigender Reihenfolge verbunden werden. Die *Phonematische Wortflüssigkeit* erfasst ebenfalls exekutive Funktionen und die verbale Flüssigkeit unabhängig semantischer Gedächtnisleistungen. Die Testperson wird aufgefordert so viele Wörter wie möglich beginnend mit dem Buchstaben S innerhalb einer Minute aufzuzählen (Rabitsch, 2014).

Die Durchführungsdauer beträgt zwischen 30 und 45 Minuten.

Auswertung: Die Summenwerte der einzelnen Untertests werden mit einem Auswertungsprogramm in normierte *z*-Werte umgerechnet. Die Rohwerte der kognitiven Funktionen werden so nach Geschlecht, Alter (49 bis 92 Jahre) und Bildungsjahren (7 bis 20 Jahre) korrigiert. Die Normwerte stammen aus einer Schweizer Stichprobe ($N = 1100$) mit einem Durchschnittsalter von 68.7 Jahren ($SD = 7.8$).

Gütekriterien: In einer umfangreichen multizentrischen Studie von Aebi (2002) wurde die diagnostische Güte der deutschsprachigen *CERADplus-NTB* mit gesunden Probanden und kognitiv leicht beeinträchtigten Alzheimerpatienten validiert. Die Klassifikationsgenauigkeit der meisten Kennwerte lag bei 80% (Pflüger, Aebi & Monsch, 2003).

Geriatrische Depressionsskala-15 (GDS-15, Sheikh & Yesavage, 1986)

Yesavage und Kollegen entwickelten 1982 mit der *Geriatrischen Depressionsskala (GDS)*, einen auf Selbstauskünften basierenden Fragebogen mit 30 leicht verständlichen, dichotomen Items zur Erfassung depressiver Symptome im höheren Lebensalter. 1986 veröffentlichten Sheikh und Yesavage die erste Kurzform mit 15 Fragen (*Geriatrische Depressionsskala-15, GDS-15*). Die *GDS-15* gilt als die am häufigsten eingesetzte Version des Verfahrens (Allgaier, Kramer, Mergl, Fejtкова & Hegerl, 2011). Zur Verbreitung der *GDS-15* haben das einfache Antwortformat und der inhaltliche Fokus auf psychologische anstelle somatischer Symptome einer Depression beigetragen. Bei älteren Menschen kann die Befragung körperlicher Symptome mit einem Verlust an Trennschärfe einhergehen (Gauggel, Schmidt & Didié, 1994). Die Auswertung erfolgt über die Bildung eines Summenwerts von maximal 15

Rohwertpunkten. Ab einem kritischen Wert ≥ 7 deutet das Screeningverfahren auf das Vorliegen eines depressiven Störungsbildes hin.

Gütekriterien: In einer Metaanalyse von Mitchell, Bird, Rizzo und Meader (2010) wurde für die *GDS-15* eine Sensitivität von 81.3 % und eine Spezifität von 78.4% ermittelt. Gauggel und Birkner (1999) fanden für die deutschsprachige Kurzform bei einem kritischen Wert ≥ 6 eine vergleichbare diagnostische Güte wie für die Langform (Sensitivität = 84.0%, Spezifität = 88.9%) und eine gute interne Konsistenz (Cronbachs $\alpha = .91$).

6.3 Soziodemographische und klinische Daten der PKB

Insgesamt wurden 12 Patienten mit Demenzverdacht in einem Zeitraum zwischen Juli 2015 und Januar 2016 untersucht. Zwei Untersuchungen fanden im Rahmen einer ambulanten Behandlung im Neurozentrum Starnberg statt. Die zehn stationären psychiatrisch-neurologischen Patienten wurden in den Räumlichkeiten der AG Neuropsychologie in der Psychiatrischen Klinik des Klinikums rechts der Isar durchgeführt. Auf Grund von Überforderung konnte bei sieben Patienten der *TMT-B* nicht durchgeführt werden, so dass dieses Verfahren nicht in die Analysen einbezogen wurde. Die Untersuchungen dauerten zwischen 65 und 120 Minuten mit einer durchschnittlichen Untersuchungsdauer von $M = 94.58$ Minuten ($SD = 18.76$). Die soziodemographischen und klinischen Informationen der Patienten werden vergleichend mit der Stichprobe der Patienten mit negativen Antwortverzerrungen (PNA) in Tabelle 12 aufgeführt.

Aus der Gegenüberstellung der beiden Stichproben PNA und PKB geht ein hoch signifikanter Altersunterschied hervor. Während die PNA im Mittel 42.20 Jahre alt waren ($SD = 9.91$, Spannweite = 22 - 52), betrug das Durchschnittsalter der PKB 74.83 Jahre ($SD = 5.95$, Spannweite = 66 - 85). In den Variablen Geschlecht, Muttersprache, Familienstand und Bildungsabschluss lagen keine statistisch bedeutsamen Differenzen vor und die Geschlechterverteilungen waren ausgeglichen. Mit Ausnahme eines Patienten waren deutsche Muttersprachler (95.5%). Neun Patienten (40.9%) befanden sich zum Untersuchungszeitpunkt in einer Partnerschaft. Acht Patienten (36.4%) hatten ein einfaches und jeweils sieben Patienten (31.8%) ein mittleres oder höheres Ausbildungsniveau absolviert.

Tabelle 12

Deskriptive Statistiken soziodemographischer und klinischer Daten der PNA und PKB, deren Gesamtstichprobe und Ergebnisse des statistischen Vergleichs der Teilstichproben.

	Total (N = 22)		PNA (n = 10)		PKB (n = 12)		p
Alter (Jahre)							
Mittelwert	60.00		42.20		74.83		.000 ^a
Standardabweichung	18.36		9.91		5.95		
Spannweite (Range)	22 - 85		22 - 52		66 - 85		
Geschlecht	n	%	n	%	n	%	
Männlich	11	50.0	6	60.0	5	41.7	.670 ^b
Weiblich	11	50.0	4	40.0	7	58.3	
Muttersprache	n	%	n	%	n	%	
Deutsch	21	95.5	9	90.0	12	100.0	.455 ^b
Andere	1	4.5	1	10.0	0	0.0	
Familienstand	n	%	n	%	n	%	
Ledig	3	13.6	2	20.0	1	8.3	1.000 ^{b,c}
Partnerschaft, nicht zusammenlebend	0	0.0	0	0.0	0	0.0	
Partnerschaft, zusammenlebend	1	4.5	1	10.0	0	0.0	
Verheiratet	8	36.4	3	30.0	5	41.7	
Verheiratet, aber getrennt	1	4.5	1	10.0	0	0.0	
Verwitwet	3	13.6	1	10.0	2	16.7	
Geschieden	6	27.3	2	20.0	4	33.3	
Schulabschluss/Berufsausbildung (BA)	n	%	n	%	n	%	
Kein Abschluss	0	0.0	0	0.0	0	0.0	.131 ^{d,e}
Hauptschulabschluss ohne BA	1	4.5	1	10.0	0	0.0	
Hauptschulabschluss mit BA	7	31.8	4	40.0	3	25.0	
Mittlere Reife mit BA	7	31.8	4	40.0	3	25.0	
Fachhochschulreife/Abitur und BA	1	4.5	0	0.0	1	8.3	
Fachhochschulabschluss	1	4.5	0	0.0	1	8.3	
Hochschulabschluss	5	22.7	1	10.0	4	33.3	
Klinische Daten	n	%	n	%	n	%	
Affektive Störungen	9	40.9	5	50.0	4	33.3	.666 ^b
Somatoforme Störungen	5	23.8	5	50.0	0	0.0	.010 ^b
Neurologische Erkrankungen	5	22.7	2	20.0	3	25.0	1.000 ^b
Psychiatrische Störungen	21	95.5	9	90.0	12	100	.455 ^b
Somatische Erkrankungen	4	18.2	2	20.0	2	16.7	1.000 ^b
Komorbidität	17	77.3	7	70.0	10	83.3	.624 ^b
Psychiatrische Medikation	15	68.2	6	60.0	9	75.0	.489 ^d

Anmerkungen.

^a t-Test für unabhängige Stichproben.

^b Exakter Test nach Fisher.

^c Berechnung des Chi-Quadrat-Tests durch Bildung von zwei Kategorien Partnerschaft ja und Partnerschaft nein.

^d Chi-Quadrat nach Pearson.

^e Berechnung des Chi-Quadrat-Tests durch Bildung von drei Kategorien einfaches, mittleres und höheres Ausbildungsniveau.

In der neurologisch-psychiatrischen Stichprobe der PKB litten alle Patienten unter mindestens einer psychiatrischen Erkrankung. Bei sieben Patienten lag eine Demenzerkrankung (Alzheimer 33.3%, Lewy-Body 25.0%) vor, die in 50% der Fälle durch bildgebende Verfahren, in 12.5% anhand eines Liquorbefundes objektiviert worden war. Darüber hinaus litten vier Patienten unter den Folgen der Einnahme psychotroper Substanzen (33.3%), zwei Patienten hatten eine kognitive Störung (16.7%). Während diese Störungsbilder bei den PNA ausgeschlossen wurden, unterschieden sich die beiden Stichproben nicht im Vorliegen affektiver Erkrankungen. Im Gegensatz zu den PNA lagen jedoch in der Gruppe der PKB keine somatoformen Störungen vor, während ein Patient (8.3%) an einem schizophrenen Störungsbild erkrankt waren. Von den PKB nahmen neun Patienten (75%) eine psychiatrische Medikation, im Sinne antidepressiver (50%), neuroleptischer (41.7%) oder antidementiver (33.3%) Medikamente, ein. Zur Behandlung somatischer Erkrankungen bekamen 11 Patienten (91.7%) regelmäßig Medikamente. Darüber hinaus lag der mittlere *GDS-15*-Summenwert zum Erhebungszeitpunkt bei 5.33 ($SD = 3.11$, Spannweite: 0 - 11). Vier Patienten (33.3%) überschritten den kritischen *GDS-15*-Wert von 6. Aus dem Screening des kognitiven Status anhand des *MMSE* ging ein durchschnittlicher Rohwert von 24.33 ($SD = 4.16$, Spannweite: 14 - 29) hervor. Der Median lag bei 26.0. Bei normorientierter Betrachtung des *MMSE* lagen 10 Patienten (83.3%) jeweils eine Standardabweichung und 9 Patienten (75.0%) jeweils zwei Standardabweichungen unter dem entsprechenden Populationsmittelwert. Ferner wurde der Zusammenhang zwischen dem kognitiven Status des *MMSE* und dem Depressionswert der *GDS-15* überprüft. Es zeigte sich ein knapp nicht statistisch bedeutsamer Zusammenhang zwischen dem *MMSE* und der *GDS-15* ($r = .552$, $p = .063$).

6.4 Statistische Datenanalysen

Für die Datenanalysen wurden die Datensätze der PNA in eine neue SPSS-Datei übertragen und die Werte der PKB ergänzt. Die Prüfung der Daten auf Vollständigkeit und Plausibilität mittels Box-Plot-Analysen zeigte, dass ein Proband Ausreißerwerte von bis zu 1.5 Interquartilen erzielt hatte. Um den Stichprobenumfang durch Ausschluss nicht zu verkleinern, wurde das Ausmaß der Ausreißer durch einen Vergleich parametrischer und non-parametrischer Verfahren ermittelt. Während sich die parametrischen und non-parametrischen Gruppenvergleiche der PVT (*t*-Tests und *U*-Tests) nicht von einander unterschieden, wichen die Korrelationskoeffizienten nach Pearson und Spearman in mehreren Variablen voneinander ab. Die Unterschiede bestanden auch nach einer Korrektur der Ausreißerwerte mittels

Winsorisierung (Tukey, 1962). Diese Korrekturmethode wurde bereits in Abschnitt 4.4.1 beschrieben und die Ausreißer wurden entsprechend der Verteilungswerte jeweils durch den nächstkleineren Wert ausgetauscht (Sachs & Hedderich, 2009). Die Analyse der Streudiagramme verdeutlichte, dass die Linearität der Zusammenhänge nicht immer gegeben war. Auf Grund unzureichender Linearität, der Ausreißerwerte und des geringen Stichprobenumfangs wurden trotz des damit einhergehenden Informationsverlustes Rangkorrelationen nach Spearman berechnet. Darüber hinaus wurde die Werteverteilungen der PKB analysiert. Die Daten der PKB waren weitgehend normalverteilt.

7 Ergebnisse: erster Teil der Studie

In diesem Kapitel werden die Ergebnisse der vorliegenden Analogstudie dargestellt. Zur Veranschaulichung werden die Daten zunächst anhand graphischer Leistungsprofile aufbereitet. Anschließend erfolgen die Ergebnisse entsprechend der Hypothesen dieser Arbeit.

7.1 Deskriptive Ergebnisdarstellung

Die Kennwerte der PVT und der kognitiven Leistungstests werden vergleichend für die Kontrollgruppe KG und die experimentellen Simulanten in zwei getrennten Leistungsprofilen in Abbildung 3 und Abbildung 4 dargestellt.

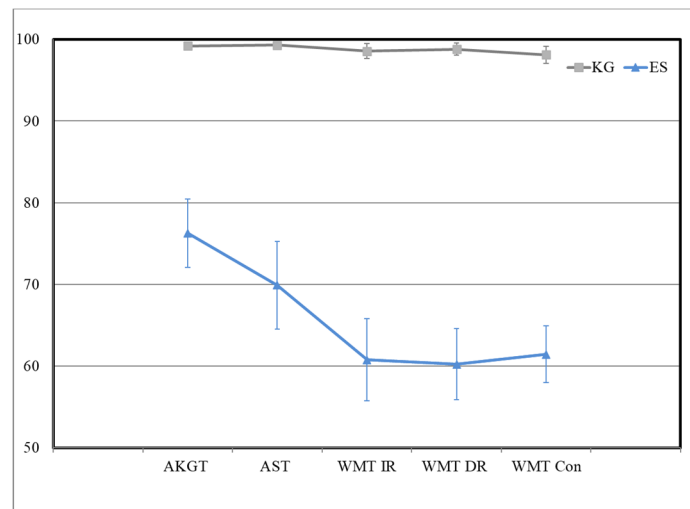


Abbildung 3. Mittlere Prozentwerte mit 95%-Konfidenzintervallen der KG ($n = 33$) und ES ($n = 33$) im *Amsterdamer Kurzzeitgedächtnistest* (AKGT), *Aggravations- und Simulationstests* (AST) und *Word Memory Test* (WMT): *unmittelbares Wiedererkennen* (WMT IR), *verzögertes Wiedererkennen* (WMT DR) und *Konsistenzmaß* (WMT Con).

Abbildung 3 verdeutlicht, dass sich die beiden experimentellen Gruppen eindeutig in ihrer Anstrengungsbereitschaft unterscheiden. Die mittleren Prozentwerte der KG, einschließlich der 95%-Konfidenzintervalle, lagen in allen drei PVT deutlich oberhalb der jeweiligen Schwellenwerte ($98\% \leq M \leq 99\%$). Dagegen rangierten die mittlere Testperformanz und die 95%-Konfidenzintervalle der ES in sämtlichen Parametern unterhalb der Schwellenwerte. Die schlechtesten Testleistungen erzielten die ES im sofortigen und verzögerten Abruf des *Word Memory Tests* (WMT IR: $M = 60.76$; WMT DR: $M = 60.23$), die beste Testperformanz im *Amsterdamer Kurzzeitgedächtnistest* (AKGT P: $M = 76.26$).

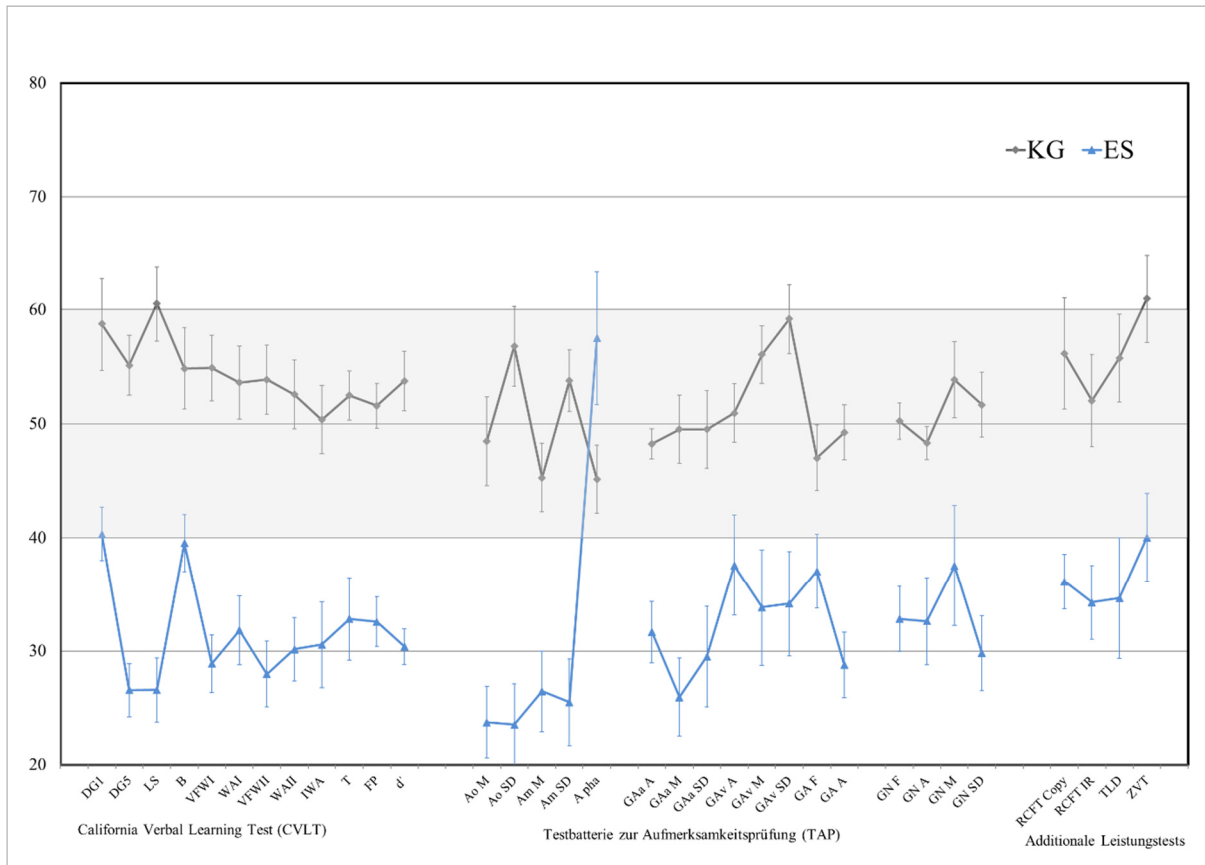


Abbildung 4. Mittlere T -Werte mit 95%-Konfidenzintervallen der KG ($n = 33$) und ES ($n = 33$): *California Verbal Learning Test (CVLT)*, *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* und *additional Performance Tests*. Graue Fläche markiert Normbereich ($40 \leq T \leq 60$).

CVLT-Untertests: *DG1*: Liste A Durchgang 1, *DG5*: Liste A Durchgang 5, *LS*: Liste A Lernsumme, *B*: Interferenzliste B, *VFWI*: Verzögerte freie Wiedergabe I, *WAI*: Wiedergabe mit Abrufhilfe I, *VFWII*: Verzögerte freie Wiedergabe II, *WAI*: Wiedergabe mit Abrufhilfe II, *IWA*: Intrusionen während Wiedergabe mit Abrufhilfe, *T*: Trefferrate bei Wiedererkennen von Liste A, *FP*: Falsch Positive bei Wiedererkennen von Liste A, *d'*: Diskriminationswert bei Wiedererkennen Liste A.

TAP-Untertests:

Alertness: *AoM*: Median RZ ohne Warnton, *AoSD*: Standardabweichung ohne Warnton, *AmM*: Median RZ mit Warnton, *AmSD*: Standardabweichung mit Warnton, *Apha*: Kennwert phasische Alertness.

Geteilte Aufmerksamkeit: *GAAa*: auditiv Auslassungen, *GAAm*: auditiv Median RZ, *GAA-SD*: auditiv Standardabweichung, *GAVa*: visuell Auslassungen, *GAVm*: visuell Median RZ, *GAV-SD*: visuell Standardabweichung, *GAF*: Fehlreaktionen gesamt, *GAA*: Auslassungen gesamt.

GoNogo: *GNF*: Fehlreaktionen, *GNA*: Auslassungen, *GNM*: Median RZ, *GNSD*: Standardabweichung.

Additional Performance Tests: *Rey Complex Figure Test and Recognition Trial RCFT Copy*: Visuokonstruktion, *RCFT IR*: kurz verzögerte visuellen Gedächtnisleistung; *Turm von London TL-D*: räumlich-strategisches Denken; *Zahlen-Verbindungs-Test ZVT*: Informationsverarbeitungsgeschwindigkeit.

Wie aus Abbildung 4 ersichtlich wird, zeigten die KG in sämtlichen mittleren T -Werten der verbalen Lern- und Merkfähigkeit im *California Verbal Learning Test (CVLT)* eine bessere Testperformance als die ES. Auch die Konfidenzintervalle überschneiden sich in keinem dieser

Parameter. In den *kurz* und *lang verzögerten freien Abrufleistungen* (*VFW I* und *VFW II*) sowie der *Anzahl erinnerter Wörter nach Durchgang 5* (*DG5*) und der *Lernsumme* (*LS*) waren die Mittelwertdifferenzen, mit knapp drei Standardabweichungen, besonders ausgeprägt. Ferner veranschaulichten die *T*-normierten Mittelwerte der *Testbatterie zur Aufmerksamkeitsprüfung* (*TAP*), dass die Testperformanz der KG in sämtlichen Aufmerksamkeitsfunktionen innerhalb des Normbereichs lag. Dagegen lag bei den Aufmerksamkeitsparametern der ES lediglich der Kennwert der *phasischen Alertness* (*Apha*) im Normbereich ($M = 57.55$), dessen Konfidenzintervall in den überdurchschnittlichen Bereich ragte. Der Kennwert der *phasischen Alertness* (*Apha*) spiegelte Leistungsschwankungen in den Reaktionszeiten für die Durchgänge mit und ohne Warnton der *Alertness* wider. Hinsichtlich der *Alertness* rangierten die Standardabweichungen (*SD*) und Mediane (*Md*) der Reaktionszeiten der ES zwischen zwei und drei Standardabweichungen unterhalb der Testperformanz der KG. Auch die Parameter der *Geteilten Aufmerksamkeit* und des *GoNogo-Paradigmas* bildeten Beeinträchtigungen unterhalb des Normbereichs ab. Lediglich die Konfidenzintervalle der *Geteilten Aufmerksamkeit Fehlreaktionen insgesamt* (*GA F*) und *visuellen Auslassungen* (*GA v A*) sowie der *Median der Reaktionszeiten* des *GoNogo-Paradigmas* (*GN Md*) ragten in den Durchschnittsbereich. Weiter zeigten sich Differenzen zwischen den beiden experimentellen Gruppen im *Turm von London* (*TL-D*), *Zahlen-Verbindungs-Test* (*ZVT*) und den beiden Subtests des *Rey Complex Figure Test and Recognition Trial* (*RCFT*), die als zusätzliche Leistungstests zusammengefasst wurden. Im *RCFT* sowie im *TL-D* lag die Testperformanz der KG innerhalb des Normbereichs, die Leistung im *ZVT* war dagegen grenzwertig überdurchschnittlich ausgeprägt ($M = 60.97$). Die ES erzielten im *ZVT* die vergleichsweise beste Testperformanz, die an den unteren Normbereich grenzte ($M = 40.03$). Dagegen präsentierten die ES im *RCFT* und *TL-D* unterdurchschnittliche Leistungen und rangierten mit einer 95%-Wahrscheinlichkeit mindestens eine *SD* unterhalb dem Normmittelwert.

7.2 Inferenzstatistische Gruppenunterschiede im Analogdesign

Im Folgenden werden die Mittelwertdifferenzen zwischen den KG und ES dargestellt. Die Gruppenunterschiede wurden anhand multipler *t*-Tests mit anschließender Bonferroni-Holm-Korrektur ermittelt und zeigten keine Abweichungen zu non-parametrischen Analysen.

7.2.1 Testperformanz in den Performanzvalidierungstests (PVT)

Tabelle 13

Deskriptive Statistiken und Effektstärken der PVT-Gruppenvergleiche.

PVT %	KG (n = 33)		ES (n = 33)		t	df	p	Cohens d
	M	SD	M	SD				
AKGT	99.19	0.93	76.26	11.76	-11.16	32	<.0005*	-2.75
AST	99.31	1.02	69.91	15.13	-11.13	32	<.0005*	-2.74
WMT IR	98.56	2.50	60.76	14.20	-15.06	34	<.0005*	-3.71
WMT DR	98.79	2.08	60.23	12.27	-17.80	34	<.0005*	-4.38
WMT Con	98.11	2.93	61.44	9.78	-20.62	38	<.0005**	-5.08

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$, ** $p < .01$.

Tabelle 13 verdeutlicht, dass die ES in allen drei PVT eine statistisch hoch signifikant schlechtere Testperformanz als die KG zeigten. Gemäß der Effektstärken waren die Gruppendifferenzen im *Word Memory Test (WMT IR, DR, Con)* stärker ausgeprägt als im *Aggravations- und Simulationstests (AST)* und *Amsterdamer Kurzzeitgedächtnistest (AKGT)*. Der größte Unterschied zeigte sich im *Konsistenzmaß des WMT (WMT Con)* mit $d = -5.08$.

7.2.2 Testperformanz in neuropsychologischen Standardverfahren

Tabelle 14

CVLT: Deskriptive Statistiken und Effektstärken der Mittelwertdifferenzen.

CVLT-Kennwert	KG (n = 33)		ES (n = 33)		t	df	p	d
	M	SD	M	SD				
Durchgang 1 (DG1)	58.76	11.51	40.33	6.62	7.97	51	<.0005*	-1.96
Durchgang 5 (DG 5)	55.12	7.43	26.55	6.59	16.53	64	<.0005**	-4.07
Lernsumme (LS)	60.55	9.29	26.58	7.96	15.95	64	<.0005**	-3.93
Interferenzliste B (B)	54.85	10.07	39.55	7.08	7.14	64	<.0005*	-1.76
Verzögerte freie Wiedergabe I (VFW I)	54.91	8.11	28.88	7.15	13.83	64	<.0005**	-3.41
Wiedergabe Abrufhilfe I (WA I)	53.61	9.06	31.82	8.53	10.06	64	<.0005**	-2.48
Verzögerte freie Wiedergabe II (VFW II)	53.88	8.51	27.97	8.17	12.62	64	<.0005**	-3.12
Wiedergabe Abrufhilfe II (WA II)	52.58	8.50	30.15	7.81	11.16	64	<.0005**	-2.75
Intrusionen Wiedergabe Abrufhilfe (IWA)	49.91	8.43	35.88	10.62	5.95	64	<.0005*	-1.46
Trefferrate (T)	51.58	5.53	32.82	10.22	9.51	52	<.0005*	-2.34
Falsch Positive (FP)	53.76	7.37	32.58	6.13	13.21	64	<.0005**	-3.25
Diskriminationswert (d')	58.76	11.51	30.39	4.43	15.62	52	<.0005**	-3.85

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$, ** $p < .01$.

In sämtlichen Parametern des *California Verbal Learning Tests (CVLT)* lagen hoch signifikante Gruppenunterschiede vor, die ES präsentierten eine schlechtere Testperformanz als die KG (Tabelle 14). Insbesondere in den Kennwerten *Durchgang 5 (DG5: $d = -4.07$)*, *Lernsumme (LS: $d = -3.93$)* und *Diskriminationsfähigkeit (d' : $d = -3.85$)* präsentierten die ES ausgeprägtere Leistungseinbußen als die KG. Die geringste Differenz wurde für die *Intrusionsneigung (IWA: $d = -1.46$)* ermittelt.

Tabelle 15

TAP: Deskriptive Statistiken und Effektstärken der Mittelwertdifferenzen.

TAP-Kennwert	KG ($n = 33$)		ES ($n = 33$)		t	df	p	d
	M	SD	M	SD				
Alertness (A)								
Median RZ ohne Warnton (Ao Md)	48.48	11.01	23.73	8.92	-10.04	64	<.0005*	-2.47
SD ohne Warnton (Ao SD)	56.82	9.86	23.52	10.07	-13.58	64	<.0005**	-3.34
Median RZ mit Warnton (Am Md)	45.27	8.47	26.45	9.97	-8.26	64	<.0005*	-2.03
SD mit Warnton (Am SD)	53.79	7.56	25.48	10.79	-12.34	64	<.0005*	-3.04
Kennwert phasische Alertness (Apha)	45.15	8.42	57.55	16.54	3.84	47	<.0005*	0.95
Geteilte Aufmerksamkeit (GA)								
auditiv Median RZ (GAa Md)	49.52	8.42	25.94	9.73	-10.52	64	<.0005**	-2.59
auditiv SD (GAa SD)	49.52	9.58	29.52	12.57	-7.27	64	<.0005*	-1.79
auditiv Auslassungen (GAa A)	48.24	3.75	31.67	7.67	-11.16	46	<.0005**	-2.75
visuell Median RZ (GAv Md)	56.06	7.09	33.42	14.16	-8.21	47	<.0005*	-2.02
visuell SD (GAv SD)	59.21	8.63	33.76	12.83	-9.45	56	<.0005**	-2.33
visuell Auslassungen (GAv A)	50.94	7.18	37.58	12.48	-5.33	51	<.0005*	-1.31
Fehlreaktionen gesamt (GA F)	47.00	8.09	37.06	9.16	-4.67	64	<.0005*	-1.15
Auslassungen gesamt (GA A)	49.24	6.83	28.79	8.16	-11.05	64	<.0005**	-2.72
Go/Nogo (GN)								
Median RZ (GN Md)	53.85	9.41	37.00	14.82	-5.51	54	<.0005*	-1.36
Standardabweichung (GN SD)	51.67	7.98	29.52	9.19	-10.46	64	<.0005*	-2.58
Fehlreaktionen (GN F)	50.24	4.48	32.82	8.07	-10.85	64	<.0005*	-2.67
Auslassungen (GN A)	48.30	4.08	32.64	10.92	-7.72	41	<.0005*	-1.90

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$, ** $p < .01$.

Tabelle 15 veranschaulicht, dass auch in den geprüften attentionalen Funktionen der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* hoch signifikante Gruppenunterschiede vorlagen. Im Untertest *Alertness* zeigten die ES in sämtlichen Durchgängen signifikant langsamere

Reaktionszeiten als die KG (*Ao Md*, *Am Md*). Mit der Verlangsamung gingen größere Standardabweichungen einher (*Ao SD*, *Am SD*). Der größte Gruppenunterschied wurde in den Standardabweichungen ohne Warnton (*Ao SD*) deutlich ($d = -3.34$). Darüber hinaus zeigten die ES kürzere Reaktionszeiten in den Durchgängen mit Warnton als ohne Warnton. Dies spiegelte sich in einem höheren Kennwert der *phasischen Alertness* (*Apha*) der ES wider, so dass dieser Kennwert als einziger Parameter oberhalb der Performanz der KG lag. Ferner zeigten sich ebenfalls hochsignifikante Gruppendifferenzen im Dual-Task-Paradigma der *geteilten Aufmerksamkeit* und im *GoNogo*-Paradigma der *Inhibitionsfähigkeit*. Neben verlangsamten *Reaktionszeiten* (*GAA Md*, *GAV Md*, *GN Md*) und größeren *Standardabweichungen* (*GAA SD*, *GAV SD*, *GN SD*) erzielten die ES signifikant häufiger *Fehlreaktionen* (*GA F*, *GN F*) und *Reaktionsauslassungen* (*GAA A*, *GAV A*, *GA A*, *GN A*). Die größten Effektstärken der *geteilten Aufmerksamkeit* lagen in den Variablen *Auslassungen* (*GA A*: $d = -2.75$) und *auditive Auslassungen* (*GAA A*: $d = -2.72$) vor. Die ausgeprägtesten Gruppendifferenzen in der *Inhibitionsfähigkeit* wurden im Kennwert *Fehlreaktionen* (*GN F*: $d = -2.67$) deutlich.

Tabelle 16

Additional Performance Tests: Descriptive Statistics and Effect Sizes of Mean Differences.

Parameter	KG ($n = 33$)		ES ($n = 33$)		t	df	p	Cohens d
	M	SD	M	SD				
RCFT								
Copy: Anzahl Richtige	56.15	13.69	36.15	6.80	-7.52	47	<.0005*	-1.85
Unmittelbares Erinnern:	52.03	11.39	34.30	9.18	-6.96	64	<.0005*	-1.71
TL-D								
Anzahl korrekter Problemlösungen	55.76	34.67	34.67	15.08	-6.52	58	<.0005*	-1.61
ZVT								
Mittlere Geschwindigkeit 4 Durchgänge	60.97	10.87	40.03	10.85	-7.83	64	<.0005*	-1.93

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$.

In den additionalen Leistungstests (Tabelle 16) unterscheiden sich die beiden Gruppen insbesondere im *Zahlen-Verbindungs-Test* (*ZVT*) ($d = -1.93$). Zudem zeigten sich auch im *Rey Complex Figure Test and Recognition Trial* (*RCFT*) und im *Turm von London* (*TL-D*) hoch signifikante Gruppenunterschiede. Die ES erzielten in sämtlichen Kennwerten eine schlechtere Testperformanz als die KG.

Zusammenfassend lagen in sämtlichen PVT und Leistungsparametern ausgeprägte Gruppenunterschiede vor. Mit Ausnahme des Kennwerts der *phasischen Alertness (Apha)* erzielten die ES durchgehend schlechtere Ergebniswerte als die KG. Die spiegelte sich auch in großen Effektstärken ($d > .80$ sensu Cohen) wider. Zur Veranschaulichung wurden die Effektstärken in Abbildung 5 dargestellt, dabei wurde auf den Kennwert *Apha* verzichtet.

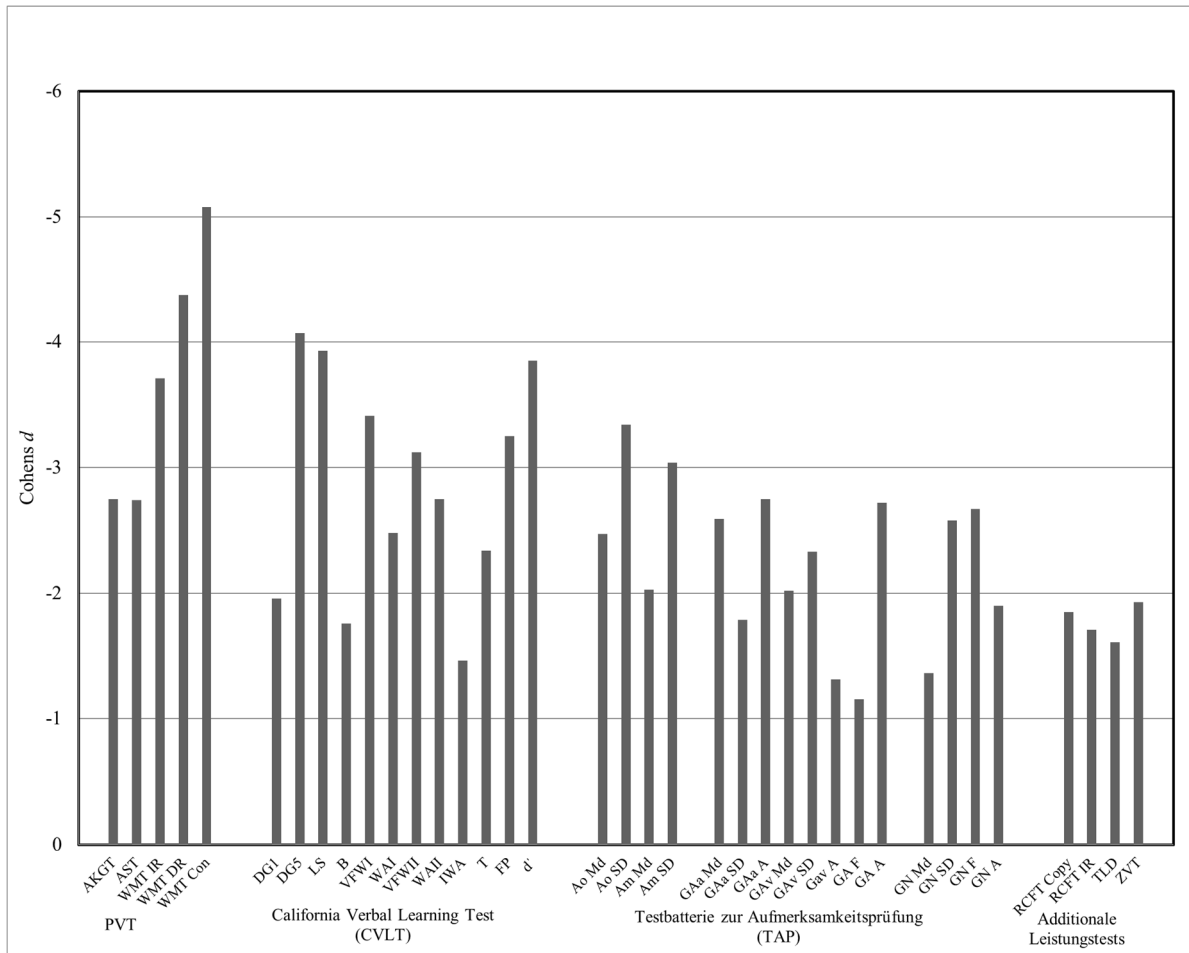


Abbildung 5. Cohens d als Maß der Effektstärke in den PVT und den neuropsychologischen Standardverfahren.

PVT: *Amsterdamer Kurzzeitgedächtnistest (AKGT)*, *Aggravations- und Simulationstests (AST)*, *Word Memory Test (WMT): unmittelbares Wiedererkennen (WMT IR)*, *verzögertes Wiedererkennen (WMT DR)*, *Konsistenzmaß (WMT Con)*.

CVLT-Untertests: *DG1:* Liste A Durchgang 1, *DG5:* Liste A Durchgang 5, *LS:* Liste A Lernsumme, *B:* Interferenzliste B, *VFWI:* Verzögerte freie Wiedergabe I, *WAI:* Wiedergabe mit Abrufhilfe I, *VFVII:* Verzögerte freie Wiedergabe II, *WAI:* Wiedergabe mit Abrufhilfe II, *IWA:* Intrusionen während Wiedergabe mit Abrufhilfe, *T:* Trefferrate bei Wiedererkennen von Liste A, *FP:* Falsch Positive bei Wiedererkennen von Liste A, *d':* Diskriminationswert bei Wiedererkennen Liste A.

TAP-Untertests:

Alertness: *AoM:* Median RZ ohne Warnton, *AoSD:* Standardabweichung ohne Warnton, *AmM:* Median RZ mit Warnton, *AmSD:* Standardabweichung mit Warnton, *Apha:* Kennwert phasische Alertness.

Geteilte Aufmerksamkeit: *GAA*: auditiv Auslassungen, *GAM*: auditiv Median RZ, *GASD*: auditiv Standardabweichung, *GVA*: visuell Auslassungen, *GM*: visuell Median RZ, *GVSD*: visuell Standardabweichung, *GAF*: Fehlreaktionen gesamt, *GAA*: Auslassungen gesamt.

GoNogo: *GNF*: Fehlreaktionen, *GNA*: Auslassungen, *GNM*: Median RZ, *GNSD*: Standardabweichung.

Additionalere Leistungstests: *Rey Complex Figure Test and Recognition Trial RCFT Copy*: Visuo-konstruktion, *RCFT IR*: kurz verzögerte visuellen Gedächtnisleistung; *Turm von London TL-D*: räumlich-strategisches Denken; *Zahlen-Verbindungs-Test ZVT*: Informationsverarbeitungsgeschwindigkeit.

Insgesamt zeigte sich die höchste Effektstärke von $d = -5.08$ im *Konsistenzmaß (WMT Con)* des *WMT*, gefolgt vom *unmittelbaren Wiedererkennen (WMT IR)* mit $d = -4.38$. Die Untertests des *WMT* wiesen größere Effektstärken als der *AKGT* und *AST* auf. In den neuropsychologischen Standardverfahren wurde die höchste Effektstärke im *CVLT* und dessen Untertest *Durchgang 5 (DG5)*: $d = -4.07$ ermittelt. Auch in den *CVLT*-Kennwerten *Lernsumme (LS)*: $d = -3.93$ und *Diskriminationsfähigkeit (d')*: $d = -3.85$ zeigten sich besonders ausgeprägte Gruppenunterschiede. Von den Kennwerten der *TAP* wiesen die *Standardabweichungen der Reaktionszeiten ohne Warnton der Alertness (AoSD)* die größte Effektstärke mit $d = -3.34$. Auch die *Standardabweichungen der Reaktionszeiten mit Warnton der Alertness (AmSD)* zeigten mit einer Effektstärke von $d = -3.04$ ausgeprägtere Gruppenunterschiede als die restlichen Kennwerte der attentionalen und additionalen Funktionen. Mit einer Effektstärke von $d = -1.15$ zeigte sich die geringste Gruppendifferenz im Kennwert *Fehlreaktionen insgesamt der geteilten Aufmerksamkeit (GA F)*.

7.2.3 Ergebnisse zu Fragestellung F1 und Hypothese H1 und H2

- F1:** Bestehen Unterschiede in der Testperformanz zwischen Personen mit maximaler und verminderter Anstrengungsbereitschaft?
- H1:** Personen mit verminderter Anstrengungsbereitschaft erzielen in den PVT-Parametern des *AKGT*, *WMT* und *AST* schlechtere Leistungen als Personen mit maximaler Anstrengungsbereitschaft.
- H2:** Personen mit verminderter Anstrengungsbereitschaft erzielen in neuropsychologischen Standardverfahren schlechtere Leistungen als Personen mit maximaler Anstrengungsbereitschaft.

Zwischen den KG und ES zeigten sich ausgeprägte Unterschiede in der Testperformanz in allen drei PVT, so dass Hypothese H1 bestätigt wurde. Ferner präsentierten die ES hypothesenkonform in allen untersuchten neuropsychologischen Standardverfahren deutliche Leistungseinbußen im Vergleich zur KG. Hypothese H2 konnte bestätigt werden.

7.3 Einfluss der Aufgabenschwierigkeit

7.3.1 Selbstbeurteilungen der subjektiven Schwierigkeit

Zur Veranschaulichung der subjektiven Schwierigkeiten erfolgt zunächst eine graphische Darstellung in Abbildung 6. Da sich keine Unterschiede zwischen den parametrischen und non-parametrischen Analysen der ordinalskalierten Selbstbeurteilungen zeigten, werden im Folgenden die Ergebnisse der parametrischen Analysen aufgeführt. Es wurden nicht nur die Unterschiede der subjektiven Schwierigkeit zwischen der KG und den ES ermittelt, sondern auch zwischen den einzelnen PVT separat für beide Gruppen. Zudem wurden die Zusammenhänge zwischen der Schwierigkeit und Testperformanz untersucht.

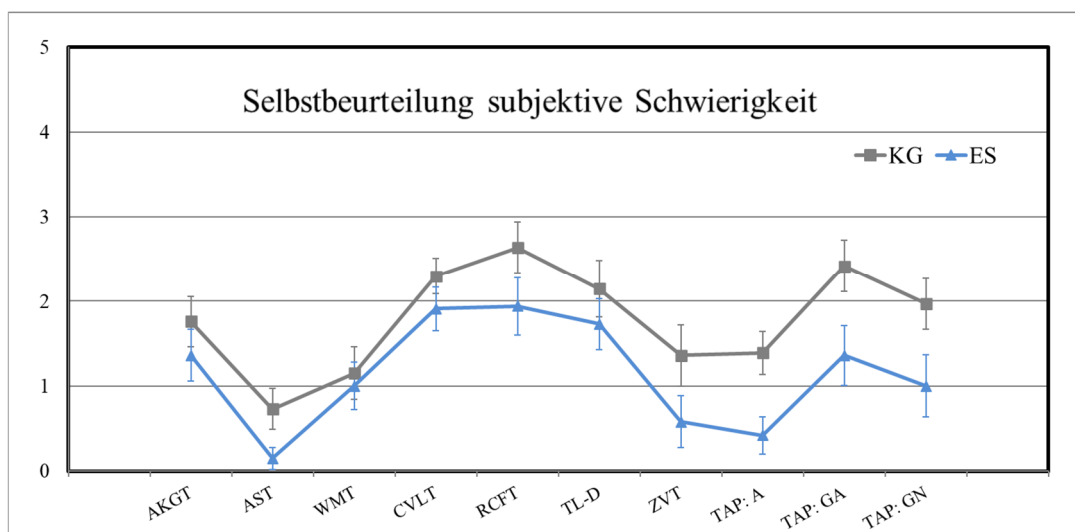


Abbildung 6. Mittelwerte mit 95%-Konfidenzintervallen der KG ($n = 33$) und ES ($n = 33$) der subjektiven Schwierigkeit (fünfstufige Einschätzung von 0 „sehr einfach“ bis 5 „sehr schwer“) im *Amsterdamer Kurzzeitgedächtnistest (AKGT)*, *Aggravations- und Simulationstests (AST)*, *Word Memory Test (WMT)*, *California Verbal Learning Test (CVLT)*, *Rey Complex Figure Test and Recognition Trial (RCFT)*, *Turm von London (TL-D)*, *Zahlen-Verbindungs-Test (ZVT)* und der *Testbatterie zur Aufmerksamkeitsprüfung (TAP): Alertness (TAP:A)*, *geteilte Aufmerksamkeit (TAP: GA)*, *GoNogo (TAP:GN)*.

Die graphische Darstellung (Abbildung 6) der subjektiven Schwierigkeiten zeigt, dass die KG sowohl die PVT als auch die Standardverfahren schwieriger einstufen als die ES. Zudem wurde der *Aggravations- und Simulationstests (AST)* von beiden Gruppen als Verfahren mit dem geringsten Schwierigkeitsgrad eingeschätzt.

Die Gruppenvergleiche der subjektiven Schwierigkeiten in Tabelle 17 verdeutlichen, dass die KG diverse Verfahren schwieriger einstufen als die ES. Dazu zählten neben dem *Aggravations- und Simulationstests (AST)* auch der *Rey Complex Figure Test and Recognition Trial (RCFT)*, der *Zahlen-Verbindungs-Test (ZVT)* und alle drei Untertests der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)*. Die größte signifikante Differenz zeigte sich in der

Beurteilung der Schwierigkeit der *Alertness* ($d = -1.48$). Dagegen lagen keine statistisch bedeutsamen Unterschiede in der subjektiven Schwierigkeit im *Turm von London (TL-D)* vor. Hervorzuheben sind zudem die Gruppendifferenzen in der Beurteilung der PVT. Während keine Gruppenunterschiede zwischen *Word Memory Test (WMT)* und *Amsterdamer Kurzzeitgedächtnistest (AKGT)* ermittelt wurden, stuften die KG den *Aggravations- und Simulationstest (AST)* schwieriger ein als die ES.

Tabelle 17

Subjektiven Schwierigkeit: Deskriptive Statistiken, Ergebnisse der t-Tests und Effektstärken.

	KG ($n = 33$)		ES ($n = 33$)		t	df	p	Cohens d
	M	SD	M	SD				
AKGT	1.76	.83	1.36	.86	-1.89	64	.063	-.47
AST	.73	.67	.15	.36	-4.32	49	< .0005**	-1.08
WMT	1.15	.87	1.00	.79	-.74	64	.462	-.18
CVLT	2.30	.59	1.91	.72	-2.43	64	.018*	-.59
RCFT	2.64	.86	1.94	.97	-3.10	64	.006**	-.76
TL-D	2.15	.94	1.73	.84	-1.93	64	.057	-.47
ZVT	1.36	1.03	.58	.87	-3.37	64	.003**	-.82
TAP: Alertness	1.39	.70	.42	.61	-5.96	64	< .0005**	-1.48
TAP: Geteilte Aufmerksamkeit	2.42	.87	1.36	.99	-4.62	64	< .0005**	-1.14
TAP: GoNogo	1.97	.85	1.00	1.03	-4.18	64	< .0005**	-1.03

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$, ** $p < .01$.

Im nächsten Auswertungsschritt wurden die Unterschiede in der subjektiven Schwierigkeit zwischen den einzelnen PVT in gruppenspezifischen Analysen untersucht. In separaten Analysen für beide Gruppen wurde die Schwierigkeit der PVT jeweils paarweise anhand von t -Tests für verbundene Stichproben miteinander verglichen.

Die Paarvergleiche der subjektiven Schwierigkeiten (Tabelle 18) verdeutlichen, dass der *AST* sowohl von den Probanden der KG als auch der ES einfacher wahrgenommen wurde als der *AKGT* und der *WMT*. Der *AKGT* wurde von beiden Gruppen schwieriger als der *WMT* eingestuft.

Tabelle 18

Paarweise Vergleiche der subjektiven Schwierigkeiten der PVT mittels t-Tests für verbundene Stichproben separat für die KG und ES.

Schwierigkeit	Gepaarte Differenzen		<i>t</i>	<i>df</i>	<i>p</i> (2-seitig)	Cohens <i>d</i>
	<i>M</i>	<i>SD</i>				
KG (<i>n</i> = 33)						
AKGT / AST	1.03	.98	6.02	32	< .0005**	1.36
AKGT / WMT	.61	1.12	3.12	32	.012*	.71
AST / WMT	-.42	1.00	-2.44	32	.042*	-.54
ES (<i>n</i> = 33)						
AKGT / AST	1.21	.89	7.80	32	< .0005**	1.80
AKGT / WMT	.36	.96	2.17	32	.042*	.44
AST / WMT	-.85	.94	-5.19	32	< .0005*	-1.41

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$, ** $p < .01$.

Des Weiteren wurde der Zusammenhang zwischen subjektiver Schwierigkeit und Testperformanz in den PVT untersucht. Dazu wurden für die Probanden der ES die Korrelationskoeffizienten nach Pearson berechnet.

Tabelle 19

*Korrelationskoeffizienten nach Pearson der ES (*n* = 33) der Testperformanz im AKGT, AST, WMT und der subjektiven Schwierigkeitsbeurteilungen.*

Testperformanz	subjektive Schwierigkeit		
	AKGT	AST	WMT
AKGT	.221	-.099	.119
AST	-.203	-.265	.205
WMT IR	-.049	-.204	-.174
WMT DR	-.015	-.200	-.064
WMT Con	-.157	.025	-.131

Aus Tabelle 19 geht hervor, dass kein statistisch bedeutsamer Zusammenhang zwischen der subjektiven Schwierigkeit und der Testperformanz in den PVT ermittelt werden konnte. Insgesamt lagen mehrheitlich negative Korrelationen zwischen der Schwierigkeit und der Anzahl an Fehlern vor, lediglich beim *AKGT* ging eine erhöhte Anzahl an Fehlern mit steigendem Schwierigkeitsgrad einher. Diese Ergebnisse waren allerdings nicht statistisch signifikant, ebenso wie die non-parametrischen Rangkorrelationen nach Spearman.

In einer weiterführenden explorativen Analyse wurden in einem Regressionsmodell die Testleistungen der PVT (Prozentwerte der Anzahl richtig wieder erkannter Items) als Kriteriumsvariable aus den drei Prädiktorvariablen Selbstbeurteilung der Schwierigkeit, Gruppenzugehörigkeit und deren Wechselwirkung (Schwierigkeit x Gruppe) vorhergesagt. Auf Grund der angenommenen Äquidistanzen wurden zur Vorhersage der PVT-Testleistungen jeweils blockweise multiple Regressionsanalysen durchgeführt (Tabelle 20).

Tabelle 20

Vorhersage der PVT-Testleistungen durch die subjektive Schwierigkeit, Gruppenzugehörigkeit und deren Wechselwirkung (Schwierigkeit x Gruppe) in der Gesamtstichprobe (N = 66) anhand blockweiser multipler Regressionsanalysen.

	<i>B</i>	<i>SE B</i>	β	<i>T</i>	<i>p</i>
AKGT					
Schwierigkeit	-.041	1.760	-.002	-.023	.981
Gruppe	-27.128	4.370	-.962	-6.208	< .0005***
WW (Schwierigkeit x Gruppe)	3.066	2.448	.197	1.253	.215
AST					
Schwierigkeit	-.232	2.755	-.008	-.084	.933
Gruppe	-27.894	3.362	-.771	-8.297	< .0005***
WW (Schwierigkeit x Gruppe)	-10.777	5.798	-.158	-1.859	.068
WMT					
Schwierigkeit	-1.063	1.483	-.043	-.716	.476
Gruppe	-36.984	2.969	-.917	-12.457	< .0005***
WW (Schwierigkeit x Gruppe)	-.854	2.206	-.032	-.387	.700

Anmerkungen. *** $p < .001$.

In allen drei PVT konnte ein bedeutsamer Anteil der Varianz der Testleistungen aufgeklärt werden: *AKGT* ($R = .82$, $R^2 = .68$, korrigiertes $R^2 = .66$, $SE = 8.27$, $df = 3$, $F = 43.36$, $p = .000$), *AST* ($R = .83$, $R^2 = .68$, korrigiertes $R^2 = .67$, $SE = 10.51$, $df = 3$, $F = 44.60$, $p = .000$) und *WMT* ($R = .94$, $R^2 = .88$, korrigiertes $R^2 = .87$, $SE = 7.30$, $df = 3$, $F = 147.05$, $p = .000$). Die genauere Betrachtung der Regressionskoeffizienten in Tabelle 20 verdeutlicht, dass in allen drei PVT weder die Schwierigkeit noch die Wechselwirkung (Schwierigkeit x Gruppe), sondern primär die experimentelle Gruppenzugehörigkeit einen statistisch hoch bedeutsamen Beitrag zur Varianzaufklärung der Testleistungen lieferte.

7.3.2 Ergebnisse zu Fragestellung F2 und Hypothese H3

F2:	Beeinflusst die subjektive Schwierigkeit eines PVT die Testperformanz?
H3:	Personen mit verminderter Anstrengungsbereitschaft zeigen in subjektiv einfach erlebten PVT weniger Fehler als in schwierigeren PVT. Es besteht ein positiver Zusammenhang zwischen der subjektiven Schwierigkeit und der Anzahl an Fehlern in PVT.

Zusammenfassend schätzten die KG die Testverfahren der vorliegenden Analogstudie insgesamt schwieriger ein als die ES. Beide Gruppen stuften den *AST* einfacher ein als den *WMT* und *AKGT*, wobei der *AKGT* als schwierigster PVT beurteilt wurde. Dennoch wurde kein Zusammenhang zwischen Schwierigkeit und Testperformanz ermittelt. Demzufolge wurde Hypothese H3 nicht bestätigt. Während die Gruppenzugehörigkeit einen erheblichen Einfluss auf die Testperformanz hatte, konnten weder die Schwierigkeit noch die Wechselwirkung Schwierigkeit x Gruppe die Leistungen in den einzelnen Verfahren vorhersagen.

7.4 Klassifikation anhand von Performanzvaliditätstests

Um die Klassifikationsstatistiken zu ermitteln, wurde zunächst die PVT-Testperformanz entsprechend der Schwellenwerte gruppiert und mit der experimentellen Gruppenzuteilung verglichen (Tabelle 21). Danach wurden die einzelnen Wertebereiche der PVT-Parameter hinsichtlich möglicher Datenseparationseffekte betrachtet, um die Voraussetzungen binärlogistischer Regressionsanalysen zu erfüllen (Tabelle 22). Ferner wurden separate binärlogistische Regressionsanalysen, ggf. mit Korrektur nach Firth, zur formal-statistischen Vorhersage der Gruppenzugehörigkeit der einzelnen Kennwerte durchgeführt. Abschließend wurden die empirischen Schwellenwerte der PVT-Kennwerte mit ROC-Analysen bestimmt.

7.4.1 Klassifikation anhand verfahrensspezifischer Schwellenwerte

Tabelle 21

Klassifikationsstatistiken gemäß verfahrensspezifischer Schwellenwerte.

PVT	Sensitivität	Spezifität	PVV	NVV	HR	OR
AKGT \leq 84 RW	97.0	100	100	97.1	98.5	Inf
AST \leq 90%	90.9	100	100	91.7	95.5	Inf
WMT IR \leq 82.5%	97.0	100	100	97.1	98.5	Inf
WMT DR \leq 82.5%	97.0	100	100	97.1	98.5	Inf
WMT Con \leq 82.5%	97.0	100	100	97.1	98.5	Inf

Aus den Klassifikationsstatistiken in Tabelle 21 geht hervor, dass die Testperformanz der KG in allen drei PVT oberhalb der kritischen Schwellenwerte der jeweiligen Testhandbücher lag. Demzufolge betrug die Spezifität und positive Vorhersagevalidität sämtlicher PVT 100%. Anhand des *AKGT* und der drei Parameter des *WMT* wurde jeweils eine Falsch-Negativ-Entscheidung erzielt (Sensitivität 97.0%). Ein Kappa-Koeffizient von $\kappa = .939$ ($p < .000$) verdeutlichte zudem eine sehr gute Interrater-Reliabilität zwischen dem *AKGT* und *WMT*. Für den *AST* wurde eine Sensitivität von 90.91% ermittelt, da die Testperformanz von drei ES im unauffälligen Wertebereich rangierte. Daraus ergab sich für den *AST* die niedrigste Klassifikationsgenauigkeit von 95.45%. Die Beobachterübereinstimmung mit dem *AKGT* lag bei $\kappa = .878$ ($p < .000$) und mit dem *WMT* bei $\kappa = .939$ ($p < .000$). Das Quotenverhältnis *OR* und damit die Chance, einen Probanden der ES auf Grund Unterschreitens der jeweiligen Schwellenwerte seiner Gruppenzugehörigkeit zuzuordnen, lief für jeden PVT-Kennwert gegen unendlich (∞).

7.4.2 Statistische Klassifikation der Gruppenzugehörigkeit

Zur Überprüfung einer möglichen Datenseparation wurden für jeden PVT-Parameter die Wertebereiche der beiden Gruppen betrachtet (Tabelle 22).

Tabelle 22

Minimale und maximale Testperformanz der einzelnen PVT-Kennwerte.

PVT-Kennwert	ES ($n = 33$)	KG ($n = 33$)	Trennwert /	Separation
	Min - Max	Min - Max	Trennbereich	
AKGT	39.00 - 86.00	87.00 - 90.00	86.00 - 87.00	**
AST	38.75 - 97.14	97.14 - 100.00	97.14	*
WMT IR	30.00 - 100.00	90.00 - 100.00		
WMT DR	32.50 - 85.00	92.50 - 100.00	85.00 - 92.50	**
WMT Con	37.50 - 85.00	87.50 - 100.00	85.00 - 87.50	**

Anmerkungen. * Quasiseparation, ** Separation.

Die minimalen und maximalen Werte des *AKGT*, *WMT DR* und *WMT Con* deuteten jeweils auf eine vollständige Datentrennung (Separation) entsprechend der Gruppenzugehörigkeit hin. Der Trennbereich des *AKGT* lag zwischen 86 und 87 Rohwertpunkten. Der Trennbereich des *WMT DR* rangierte zwischen 85% und 92.5%, der Trennbereich des *WMT Con* zwischen 85% und 87.5%. Eine quasi-vollständige Trennung der Daten (Quasiseparation) ergab sich für den Prozentwert richtiger Antworten im *AST* mit einem Trennwert von 97.14%. Mit dem *WMT IR* wurde dagegen keine Separation der Daten erzielt. Aus Tabelle 22 geht eine Klassifikationsgenauigkeit von 100% für den *AKGT*, *WMT DR* und *WMT Con* hervor, so dass

weitere Analysen mit diesen Kennwerten redundant erschienen. Zur formal-statistischen Vorhersage der Gruppenzugehörigkeit mittels *AST* und *WMT IR* wurde jeweils eine binär-logistische Regressionsanalyse durchgeführt (Tabelle 23).

Tabelle 23

Separate binär-logistische Regressionsanalysen jeweils mit dem AST und WMT IR.

Kennwert	<i>B</i>	<i>SE B</i>	Wald	<i>df</i>	<i>p</i>	<i>OR</i>	Klassifikation <i>n</i> (%)		
							total	ES	KG
AST	-9.55	455.94	.000	1	.983	.000	65 (98.5)	32 (97.0)	33 (100)
Konstante	926.08	44289.98	.000	1	.983	.			
WMT IR	-.309	.104	8.884	1	.003	.734	65 (98.5)	32 (97.0)	33 (100)
Konstante	26.83	9.65	7.728	1	.005	4.47E+11			

Aus den logistischen Regressionsanalysen in Tabelle 23 geht hervor, dass anhand des *AST* 98.5% der Probanden korrekt klassifiziert konnten. Ein ES wurde fälschlicherweise der KG zugeordnet. Laut der Schätzung von R^2 nach Nagelkerke lag der Anteil aufgeklärter Varianz bei 97.4% ($\chi^2(1) = 86.49$, $p \leq .000$, Nagelkerke's $R^2 = .974$). Neben den sehr guten Klassifikationsgenauigkeiten deuteten auch die hohen Regressionskoeffizienten und Standardfehler der Regressionsgleichung auf eine Separation bzw. Quasiseparation der Daten hin. Um die Verletzung fehlender Prädiktorenkombinationen zu berücksichtigen, wurde die Korrektur nach Firth zur Penalisation der *M-L*-Funktion angewandt. Dadurch konnten die Schwierigkeiten der Parameterbestimmung im Regressionsmodell statistisch korrigiert werden. Nach dem logistischen Regressionsmodell mit Korrektur nach Firth lag die Vorhersagegenauigkeit weiterhin bei 98.5% mit einer Falsch-Negativ-Fehlklassifikation.

$$p(ES) = \frac{e [102.35 - 1.07 (AST)]}{1 + e [102.35 - 1.07 (AST)]}$$

Auch mit dem *WMT IR* konnte die experimentellen Gruppenzugehörigkeit sehr gut vorhergesagt werden ($\chi^2(1) = 86.49$, $p \leq .000$, Nagelkerke's $R^2 = .974$). Die Varianzaufklärung R^2 nach Nagelkerke betrug wie beim *AST* 97.4%. Die Gruppenzuteilung entsprach ebenfalls zu 98.5% den experimentellen Bedingungen, mit einer Fehlklassifikation eines ES als Proband der KG.

7.4.3 Ermittlung empirischer Schwellenwerte

Tabelle 24

ROC-Kurven der PVT: AUC und Wilcoxon-Teststatistiken.

Kennwert	AUC	SE	p asymptotisch	CI (95%)	
				Untergrenze	Obergrenze
AKGT	1.000	0.000	< .0005***	1.000	1.000
AST	0.998	0.003	< .0005***	0.993	1.000
WMT IR	0.980	0.021	< .0005***	0.939	1.000
WMT DR	1.000	0.000	< .0005***	1.000	1.000
WMT Con	1.000	0.000	< .0005***	1.000	1.000

Anmerkungen. *** $p < .001$.

Die ROC-Kurven des *AKGT*, *WMT DR* und *WMT Con* verliefen jeweils exakt entlang der Begrenzungslinien des Quadrats, was eine perfekte Gruppendifferenzierung verdeutlichte. Die ROC-Kurven des *WMT IR* und *AST* näherten sich jeweils den Begrenzungslinien an und gelten nach Hosmer und Lemeshow (2000) als herausragend. Die optimalen Schwellenwerte werden in Tabelle 25 zusammengefasst.

Tabelle 25

Klassifikationsstatistiken des AGKT, AST und WMT.

PVT	Schwellenwert	Sensitivität	Spezifität	PVV	NVV	HR	OR	J
AKGT	86.5	100	100	100	100	100	Inf	100
AST	95.7	97.0	100	100	97.1	98.5	Inf	97.0
WMT IR	83.8	97.0	100	100	97.1	98.5	Inf	97.0
	91.3	97.0	100	100	97.1	98.5	Inf	97.0
WMT DR	88.8	100	100	100	100	100	Inf	100
WMT Con	86.3	100	100	100	100	100	Inf	100

Tabelle 25 verdeutlicht, dass die diagnostische Güte des *AST*, *AKGT* und *WMT* durch Heraufsetzen der Schwellenwerte weiter verbessert werden konnte. Bei einem Schwellenwert von 86.5 wurde für den *AKGT* eine perfekte Klassifikationsgenauigkeit von 100% ermittelt, ebenso für die Kennwerte *WMT DR* (Schwellenwert 88.75%) und *WMT Con* (Schwellenwert 86.25%). Mit einem Schwellenwert von 95.72% wurde die Sensitivität des *AST* auf 97.0% und die Klassifikationsgenauigkeit auf 98.5% verbessert. Die ROC-Kurven ergaben für den *WMT IR* jeweils zwei optimale Schwellenwerte von 83.75% und 91.25% mit jeweils einer Klassifikationsgenauigkeit von 98.5%. Demzufolge konnte durch einen konservativeren

Schwellenwert die Entscheidungsrate nicht verbessert werden. Die *PVV* des *AKGT*, *WMT IR* und *WMT Con* betrug jeweils 100%. Gemessen an *OR*, lief die Chance, anhand der optimalen Schwellenwerte des *AKGT*, *AST* und *WMT* negative Antwortverzerrungen in der Gruppe der ES zu erfassen, gegen unendlich. Insgesamt lagen (fast) perfekte Klassifikationsstatistiken für den *AKGT* und *WMT* vor. Mit einer minimal geringeren Sensitivität zeichnete sich auch der *AST* als zuverlässiger PVT aus.

7.4.4 Ergebnisse zu Fragestellung F3 und Hypothese H4

F3: Lassen sich die Klassifikationsstatistiken der Autoren des *AST* replizieren?
H4: Sämtliche Probanden der experimentellen Analogstudie können anhand des *AST* entsprechend ihrer Gruppenzugehörigkeit korrekt klassifiziert werden.

95.5% der Probanden konnten mit dem *AST* und einem Schwellenwert von 90% korrekt klassifiziert werden. Mit einem Schwellenwert von 95.72% konnte die Klassifikationsgenauigkeit des *AST* auf 98.5% erhöht werden, die mit einer Verbesserung der Sensitivität von 97% einherging. Allerdings konnte die Klassifikationsgenauigkeit von 100% der Testautoren Eberl und Wilhelm (2007) und die fehlerfreie Vorhersage der Gruppenzugehörigkeit mit dem *WMT* und *AKGT* nicht erzielt werden. Somit kann Hypothese H4, trotz hervorragender Klassifikationsstatistiken des *AST*, nicht als bestätigt gelten.

7.5 Klassifikation anhand neuropsychologischer Standardverfahren

7.5.1 Klassifikation anhand von CVLT-Kennwerten

Nach Überprüfung der Wertebereiche der Parameter des *California Verbal Learning Test (CVLT)* konnte eine (Quasi-) Separation der Daten ausgeschlossen werden (siehe Tabelle A-1). Anschließend wurde die Gruppenzugehörigkeit mit einer schrittweisen binär-logistischen Regressionsanalyse mit Rückwärts- und Vorwärtsselektion (Tabelle 26) vorhergesagt. Da die Gültigkeit der Regressionsmodelle durch das Fehlen leerer Zellen eingeschränkt war, wurde für beide Modelle eine Korrektur nach Firth berechnet. Ferner wurden Schwellenwerte aus ROC-Kurven und Klassifikationsstatistiken bestimmt.

Tabelle 26

CVLT: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.

	Kennwert	B	SE B	Wal	df	p	OR	Klassifikation n (%)		
								total	ES	KG
rückwärts	LS	-20.08	559.26	.001	1	.971	.000	66 (100)	33 (100)	33 (100)
Schritt 11	FP	-10.94	309.29	.001	1	.972	.000			
	Konstante	1266.46	35098.5	.001	1	.971	.			
vorwärts	DG5	-1.49	1.76	.713	1	.399	.227	65 (98.5)	33 (100)	32 (97.0)
Schritt 2	FP	-1.62	2.01	.652	1	.419	.198			
	Konstante	118.66	143.29	.686	1	.408	3.412E+51			

CVLT: LS: Lernsumme, FP: Falsch Positive, DG5: Durchgang 5.

Die schrittweise Regressionsanalyse mit Rückwärtsselektion (Tabelle 26) zeigt, dass sämtliche Probanden beider Gruppen anhand der Parameter des *CVLT* korrekt klassifiziert werden konnten ($\chi^2(2) = 91.50$, $p \leq .000$, Nagelkerke's $R^2 = 1.0$). Nach Schritt 11 der schrittweisen Prädiktorselektion konnte mit den beiden Kennwerten *Lernsumme (LS)* und *Falsch-Positive (FP)* weiterhin die Gruppenzugehörigkeit sämtlicher Probanden korrekt vorhergesagt werden. Auf Grund fehlender Zellen wurde die Gruppenvorhersage erneut mit den korrigierten Koeffizienten nach Firth berechnet. Dadurch sank die Klassifikationsgenauigkeit mit einer Falsch-Positiv-Fehlklassifikation auf 98.5%. In dem korrigierten Modell war die Vorhersage mit dem Prädiktor *LS* statisch bedeutsam ($\chi^2 = 18.96$, $p \leq .000$), im Gegensatz zum Prädiktor *FP* ($\chi^2 = 3.01$, $p \leq .083$).

$$p(ES) \text{ rückwärts} = \frac{e [16.59 - .21 (LS) - .19 (FP)]}{1 + e [16.59 - .21 (LS) - .19 (FP)]}$$

In Schritt 2 der Regressionsanalyse mit Vorwärtsselektion konnte das Vorhersagemodell durch die Aufnahme des Kennwerts *Falsch-Positive (FP)* zum ersten Prädiktor *Durchgang 5 (DG5)* verbessert werden. Dadurch stieg die Klassifikationsgenauigkeit von 92.4% auf 98.5% mit einer Falsch-Positiv-Entscheidung. Die Varianzaufklärung gemäß der Schätzung nach Nagelkerke lag bei 97.9% ($\chi^2(2) = 87.435$, $p \leq .000$, Nagelkerke's $R^2 = .979$). Die korrigierte Gruppenvorhersage nach Firth lag ebenfalls bei 98.5% für die beiden signifikanten Prädiktoren *DG5* ($\chi^2 = 18.83$, $p \leq .000$) und *FP* ($\chi^2 = 6.67$, $p \leq .010$).

$$p(ES) \text{ vorwärts} = \frac{e [18.58 - .25(DG5) - .23(FP)]}{1 + e [18.58 - 0.25(DG5) - .23(FP)]}$$

Für beide korrigierten Vorhersagemodelle wurden jeweils die gleichen sehr guten Übereinstimmungen mit dem *WMT* ($\kappa = .939$, $p < .000$) und dem *AKGT* ($\kappa = .939$, $p < .000$) ermittelt.

Tabelle 27

ROC-Kurven des CVLT: AUC und Wilcoxon-Teststatistiken.

CVLT-Kennwert	AUC	SE	p asymptotisch	CI (95%)	
				Untergrenze	Obergrenze
Durchgang 1 (DG1)	.931	.029	< .0005***	.875	.987
Durchgang 5 (DG 5)	.992	.007	< .0005***	.978	1.000
Lernsumme (LS)	.998	.003	< .0005***	.992	1.000
Interferenzliste B (B)	.892	.040	< .0005***	.814	.970
Verzögerte freie Wiedergabe I (VFW I)	.991	.009	< .0005***	.974	1.000
Wiedergabe mit Abrufhilfe I (WA I)	.953	.022	< .0005***	.909	.996
Verzögerte freie Wiedergabe II (VFW II)	.974	.015	< .0005***	.944	1.000
Wiedergabe mit Abrufhilfe II (WA II)	.964	.019	< .0005***	.927	1.000
Intrusionen Wiedergabe mit Abrufhilfe (IWA)	.845	.050	< .0005***	.747	.944
Trefferrate (T)	.921	.038	< .0005***	.848	.995
Falsch Positive (FP)	.965	.022	< .0005***	.923	1.000
Diskriminationswert (d')	.990	.008	< .0005***	.976	1.000
p (ES) Lernsumme (LS), Falsch Positive (FP)	.999	.002	< .0005***	.996	1.000
p (ES) Durchgang 5 (DG 5), Falsch Positive (FP)	.999	.002	< .0005***	.996	1.000

Anmerkungen. *** $p < .001$.

Die ROC-Kurvenverläufe sämtlicher Kennwerte des *CVLT* näherten sich der linken oberen Ecke an. Die Kurven der beiden Regressionsmodelle wiesen die größten Überschneidungen mit den Bezugslinien auf. Dies spiegelte sich auch in den höchsten AUC-Werten von jeweils .999 wider. Die AUC-Werte der separaten *CVLT*-Kennwerte rangierten zwischen .998 (*Lernsumme LS*) und .845 (*Intrusionen der Wiedergabe mit Abrufhilfe IWA*). Gemäß Hosmer und Lemeshow (2000) konnten die AUC-Werte allesamt als ausgezeichnet bis herausragend eingestuft werden.

Die optimalen Schwellenwerte der *CVLT*-Parameter, die aus den ROC-Analysen durch die Maximierung der Sensitivität und Spezifität ermittelt wurden, werden in Tabelle 28 aufgeführt. Eine vollständige Auflistung der Schwellenwerte ist in Tabelle A-2 abgebildet.

Tabelle 28

Klassifikationsstatistiken der CVLT-Parameter.

CVLT-Kennwert	Cut-Off	Sensitivität	Spezifität	PVV	NVV	HR	OR	J
Durchgang 1 (DG1)	44.5	75.8	93.9	92.6	79.5	84.8	48.2	69.7
Durchgang 5 (DG 5)	36.0	93.9	100	100	94.3	96.9	Inf	93.9
Lernsumme (LS)	42.5	97.0	97.0	97.0	97.0	97.0	1045.4	94.0
Interferenzliste B (B)	44.5	69.7	93.9	92.0	75.6	81.8	35.4	63.6
Verzögerte freie Wiedergabe I (VFW I)	41.5	97.0	97.0	97.0	97.0	97.0	1045.4	94.0
Wiedergabe mit Abrufhilfe I (WA I)	43.0	90.9	84.8	85.7	90.3	87.8	55.7	75.7
Verzögerte freie Wiedergabe II (VFW II)	33.5	84.8	100	100	86.8	92.4	Inf	84.8
Wiedergabe mit Abrufhilfe II (WA II)	39.5	87.9	93.9	93.5	88.6	90.0	111.8	81.8
Intrusionen Wiedergabe Abrufhilfe (IWA)	46.5	87.9	75.8	78.4	86.2	81.9	22.7	63.7
Trefferrate (T)	39.0	84.8	93.9	93.3	86.1	89.3	85.8	78.7
Falsch Positive (FP)	34.5	72.7	100	100	78.6	86.4	Inf	72.7
Diskriminationswert (d')	32.5	87.9	100	100	89.2	94.0	Inf	87.9
p (ES) LS, FP	.405	100	97.0	97.1	100	98.5	Inf	97.0
	.623	97.0	100	100	97.1	98.5	Inf	97.0
p (ES) DG5, FP	.477	100	97.0	97.1	100	98.5	Inf	97.0
	.656	97.0	100	100	97.1	98.5	Inf	97.0

Die Klassifikationsstatistiken des *CVLT* in Tabelle 28 verdeutlichen, dass mit den beiden Vorhersagemodellen die höchsten Klassifikationsgenauigkeiten von jeweils 98.5% erreicht wurden. Für beide Vorhersagemodelle konnten je zwei optimale Schwellenwerte ermittelt werden. Bei unveränderter Klassifikationsgenauigkeit von 98.5% konnte durch Heraufsetzen des Schwellenwerts jeweils eine Falsch-Positiv-Entscheidung zu Lasten einer Falsch-Negativ-Fehlklassifikation verhindert werden und umgekehrt.

Darüber hinaus wird ersichtlich, dass mit den Schwellenwerten der einzelnen Variablen *Durchgang 5 (DG5: $T = 36.0$)*, *Lernsumme (LS: $T = 42.5$)* und *verzögerte freie Wiedergabe I (VFW I: $T = 41.5$)* jeweils eine Klassifikationsgenauigkeit von 97% erzielt wurde. Ferner wurden mit den einzelnen Parametern *Durchgang 5 (DG5)*, *verzögerte freie Wiedergabe II (VFW II)*, *Falsch-Positive (FP)* und *Diskriminationswert (d')* keine Falsch-Positiv-Fehlklassifizierungen getroffen. Die *PVV* lag bei 100%, die Wahrscheinlichkeit negativer Antwortverzerrungen bei einer Testperformanz unterhalb der Schwellenwerte ging gegen unendlich. Lediglich die Kennwerte *Wiedergabe mit Abrufhilfe I (WA I)* und *IWA* entsprachen mit einer Spezifität von 84.8% bzw. 75.8% nicht den BV-Konventionen. Die Betrachtung der

restlichen *CVLT*-Kennwerte zeigte ein homogenes Bild hinsichtlich der Klassifikationsgenauigkeiten, die zwischen 81.8% und 92.4% rangierten.

7.5.2 Ergebnisse zu Fragestellung F4 und Hypothese H5

- F4:** Inwiefern eignen sich verbal-mnestische und attentionale Testvariablen neuropsychologischer Standardverfahren zur Vorhersage der Gruppenzugehörigkeit?
- H5:** Die experimentelle Gruppenzugehörigkeit kann zuverlässig aus den *CVLT*-Parametern vorhergesagt werden.

Aus den Ergebnissen geht hypothesenkonform hervor, dass die experimentelle Gruppenzugehörigkeit anhand der *CVLT*-Kennwerte vorhergesagt und Hypothese H5 somit bestätigt werden konnte. Anhand der Prädiktorkombinationen *Lernsumme (LS)* und *Falsch-Positive (FP)* sowie *Durchgang 5 (DG5)* und *Falsch-Positive (FP)* konnten jeweils 98.5% der Probanden korrekt klassifiziert werden, wobei je ein Proband der KG fälschlicherweise den ES zugeordnet wurde. Darüber hinaus wurden für die Schwellenwerte der Parameter *Durchgang 5 (DG5)*, *Lernsumme (LS)*, *verzögerte freie Wiedergabe I (VFW I)* und *Diskriminationswert (d')* Klassifikationsgenauigkeiten zwischen 94 und 97% ermittelt.

7.5.3 Klassifikation anhand von TAP-Kennwerten

Nachdem eine Datenseparation der Kennwerte der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* ausgeschlossen wurde (Tabelle A-3), wurden statistisch bedeutsame Prädiktoren der *TAP* anhand schrittweiser binär-logistische Regressionsanalysen mit Rückwärts- und Vorwärtsselektion ermittelt. Anschließend wurden ROC-Kurvenverläufe und Klassifikationsstatistiken der *TAP*-Kennwerte analysiert. Zudem wurde in weiteren Regressionsanalysen die Gruppenzugehörigkeit aus einer Kombination aus Prädiktoren des *CVLT* und der *TAP* vorhergesagt. Anschließend wurden Schwellenwerte aus ROC-Kurven und Klassifikationsstatistiken abgeleitet.

Alertness

Die schrittweise logistische Regressionsanalyse mit Rückwärtsselektion in Tabelle 29 verdeutlicht, dass nach Schritt 5 mit dem Kennwert *Ao SD (Standardabweichungen der Reaktionszeiten ohne Warnton)* 92.4% der Probanden korrekt vorhergesagt werden konnten. Drei ES wurden als Falsch-Negativ-Klassifikationen den KG zugeordnet, zwei KG wurden als ES fehlklassifiziert. Auch die Regressionsanalyse mit Vorwärtsselektion ergab, dass bereits nach Schritt 1 mit dem Prädiktor *Ao SD* das Vorhersagemodell durch die Aufnahme

weiterer Prädiktoren nicht mehr verbessert werden konnte. Demzufolge resultierte aus beiden Analysen das gleiche Vorhersagemodell, mit dem 83.5% der Gesamtvarianz nach Schätzung gemäß Nagelkerke aufgeklärt wurde ($\chi^2(1) = 64.919, p \leq .000$, Nagelkerke's $R^2 = .835$). Die Übereinstimmung mit dem *WMT* ($\kappa = .879, p < .000$) und dem *AKGT* ($\kappa = .818, p < .000$) war jeweils sehr gut.

Tabelle 29

TAP Alertness: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.

	Kennwert	<i>B</i>	SE <i>B</i>	Wald	<i>df</i>	<i>p</i>	<i>OR</i>	Klassifikation <i>n</i> (%)		
								total	ES	KG
rückwärts	Ao SD	-.19	.04	18.943	1	.000	.831	61 (92.4)	30 (90.9)	31 (93.9)
Schritt 5	Konstante	7.31	1.78	16.892	1	.785	1495.417			
vorwärts	Ao SD	-.19	.04	18.943	1	.000	.831	61 (92.4)	30 (90.9)	31 (93.9)
Schritt 1	Konstante	7.31	1.78	16.892	1	.785	1495.417			

Alertness: AoSD: Standardabweichung ohne Warnton

Geteilte Aufmerksamkeit

Mit der logistischen Regressionsanalyse mit Rückwärtsselektion in Tabelle 30 konnten sämtliche Probanden korrekt klassifiziert werden. Nach Schritt 2 und Ausschluss des Prädiktors *Gav A (visuelle Auslassungen)* wurden keine weiteren Variablen aus dem Modell entfernt, da dies eine signifikante Verschlechterung der Vorhersage zur Folge hätte. Mit dem Vorhersagemodell konnte die Gesamtvarianz nach Schätzung gemäß Nagelkerke vollständig aufgeklärt werden ($\chi^2(7) = 91.495, p \leq .000$, Nagelkerke's $R^2 = 1.00$). Die korrigierte Gruppenvorhersage nach Firth verdeutlichte, dass nur der Prädiktor *GAv SD* (Standardabweichungen der visuellen Reaktionszeiten) einen signifikanten Einfluss auf das Modell hatte ($\chi^2 = 5.12, p \leq .024$). Die anderen Prädiktoren konnten nicht gegen den Zufall abgesichert werden ($p > .05$). Die Klassifikationsgenauigkeit lag bei 97.0%.

$$p(ES)_{rück} = \frac{e [22.60 - .29(GAa A) - .17(GAa Md) - .01(GAa SD) + .10(GAv Md) - .31 (GAv SD) + .12 (GA F) + .03 (GA A)]}{1 + e [22.60 - .29(GAa A) - .17(GAa Md) - .01(GAa SD) + .10(GAv Md) - .31 (GAv SD) + .12 (GA F) + .03 (GA A)]}$$

Für das Vorhersagemodell der Rückwärtsselektion wurde eine sehr gute Interrater-Reliabilität mit dem *AKGT* und *WMT* von jeweils $\kappa = .909 (p < .000)$ ermittelt.

Tabelle 30

TAP Geteilte Aufmerksamkeit: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.

	Kennwert	B	SE B	Wald	df	p	OR	Klassifikation n (%)		
								total	ES	KG
rückwärts	GAa A	-12.86	390.45	.001	1	.974	.000	66 (100)	33 (100)	33 (100)
Schritt 2	GAa Md	-8.95	278.56	.001	1	.974	.000			
	GAa SD	-3.21	121.71	.001	1	.979	.040			
	GAv Md	3.66	147.73	.001	1	.980	38649			
	GAv SD	-17.98	529.51	.001	1	.973	.000			
	GA F	6.59	230.75	.001	1	.977	725.662			
	GA A	3.04	117.65	.001	1	.979	20.908			
	Konstante	1313.41	38046.17	.001	1	.972	.			
vorwärts	GAa A	-.27	.12	4.868	1	.027	.761	62 (93.9)	31 (93.9)	31 (93.9)
Schritt 3	GAa Md	-.20	.11	3.391	1	.066	.822			
	GAv SD	-.28	.14	4.013	1	.045	.750			
	Konstante	33.27	14.37	5.361	1	.021	2.821E+14			

GAaA: auditiv Auslassungen, *GAaMd*: auditiv Median RZ, *GAaSD*: auditiv Standardabweichung, *GAvA*: visuell Auslassungen, *GAvMd*: visuell Median RZ, *GAvSD*: Standardabweichung, *GA F*: Fehlreaktionen gesamt, *GA A*: Auslassungen gesamt.

Mit der logistischen Regressionsanalyse mit Rückwärtsselektion in Tabelle 30 konnten sämtliche Probanden korrekt klassifiziert werden. Nach Schritt 2 und Ausschluss des Prädiktors *Gav A* (*visuelle Auslassungen*) wurden keine weiteren Variablen aus dem Modell entfernt, da dies eine signifikante Verschlechterung der Vorhersage zur Folge hätte. Mit dem Vorhersagemodell konnte die Gesamtvarianz nach Schätzung gemäß Nagelkerke vollständig aufgeklärt werden ($\chi^2(7) = 91.495$, $p \leq .000$, Nagelkerke's $R^2 = 1.00$). Die korrigierte Gruppenvorhersage nach Firth verdeutlichte, dass nur der Prädiktor *GAv SD* (Standardabweichungen der visuellen Reaktionszeiten) einen signifikanten Einfluss auf das Modell hatte ($\chi^2 = 5.12$, $p \leq .024$). Die anderen Prädiktoren konnten nicht gegen den Zufall abgesichert werden ($p > .05$). Die Klassifikationsgenauigkeit lag bei 97.0%.

$$p(ES)_{\text{rück}} = \frac{e [22.60 - .29(GAa A) - .17(GAa Md) - .01(GAa SD) + .10(GAv Md) - .31 (GAv SD) + .12 (GA F) + .03 (GA A)]}{1 + e [22.60 - .29(GAa A) - .17(GAa Md) - .01(GAa SD) + .10(GAv Md) - .31 (GAv SD) + .12 (GA F) + .03 (GA A)]}$$

Für das Vorhersagemodell der Rückwärtsselektion wurde eine sehr gute Interrater-Reliabilität mit dem *AKGT* und *WMT* von jeweils $\kappa = .909$ ($p < .000$) ermittelt.

In das Vorhersagemodell der logistischen Regressionsanalyse mit Vorwärtsselektion wurde in Schritt 1 zunächst der Prädiktor *auditive Auslassungen (GAa A)* aufgenommen. Dadurch konnten 75.7% der Gesamtvarianz aufgeklärt werden. In zwei weiteren Schritten wurden die beiden Prädiktoren *Median der Reaktionszeiten auditiv (GAa Md)* und *Standardabweichungen der Reaktionszeiten visuell (GAv SD)* in das Modell aufgenommen. Nach Schritt 3 konnte die Gruppenzugehörigkeit von 93.9% der Probanden korrekt vorhergesagt werden. Sowohl das Vorhersagemodell insgesamt ($\chi^2(3) = 78.071, p \leq .000$, Nagelkerke's $R^2 = .925$) als auch der Einfluss der beiden Prädiktoren *GAa A* und *GAv SD* konnte als statistisch signifikant abgesichert werden. Die Übereinstimmung des Vorhersagemodells mit dem *WMT* und dem *AKGT* betrug jeweils $\kappa = .848$ ($p < .000$).

Go/Nogo

Tabelle 31

TAP Go/Nogo: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.

	Kennwert	B	SE B	Wald	df	p	OR	Klassifikation n (%)		
								total	ES	KG
rückwärts	GN F	-.30	.08	14.291	1	.000	.742	61 (92.4)	31 (93.9)	30 (90.9)
Schritt 3	GN Md	-.15	.06	7.038	1	.008	.864			
	Konstante	19.59	5.35	13.401	1	.000	321288385.7			
vorwärts	GN F	-.20	.07	7.959	1	.005	.822	61 (92.4)	31 (93.9)	30 (90.9)
Schritt 2	GN SD	-.15	.06	7.412	1	.006	.859			
	Konstante	14.70	3.88	14.384	1	.000	2429196.96			

GN F: GoNogo Fehlreaktionen, *GN A*: GoNogo Auslassungen, *GN M*: GoNogo Median RZ, *GN SD*: GoNogo Standardabweichung.

Die logistische Regressionsanalyse mit Rückwärtsselektion in Tabelle 31 verdeutlicht, dass nach Schritt 3 und Ausschluss der beiden Variablen *Auslassungen (GN A)* und *Standardabweichungen der Reaktionszeiten (GN SD)* 92.4% der Probanden mit den beiden Prädiktoren *Fehlreaktionen (GN F)* und *Median der Reaktionszeiten (GN Md)* korrekt vorhergesagt werden konnten. Sowohl das Vorhersagemodell ($\chi^2(2) = 67.286, p \leq .000$, Nagelkerke's $R^2 = .852$) als auch die beiden Prädiktoren waren statistisch signifikant. Mit dem *WMT* ($\kappa = .879, p < .000$) und dem *AKGT* ($\kappa = .818, p < .000$) wurde jeweils eine sehr hohe Interrater-Reliabilität für das Vorhersagemodell ermittelt.

In Schritt 1 der logistischen Regressionsanalyse mit Vorwärtsselektion konnten 89.4% der Probanden mit dem Prädiktor *Fehlreaktionen (GN F)* korrekt vorhergesagt werden und 75.1%

der Gesamtvarianz aufgeklärt werden. Durch Aufnahme des Prädiktors *Standardabweichung der Reaktionszeiten (GN SD)* in Schritt 2 verbesserte sich die Vorhersage auf 92.4%. Das Vorhersagemodell ($\chi^2(2) = 65.480$, $p \leq .000$, Nagelkerke's $R^2 = .839$) und die beiden Prädiktoren *GN F* und *GN SD* waren statistisch signifikant. Der Kappa-Koeffizient betrug mit dem *WMT* $\kappa = .879$ ($p < .000$), mit dem *AKGT* $\kappa = .818$ ($p < .000$).

Tabelle 32

ROC-Kurven der TAP-Kennwerte: AUC und Wilcoxon-Teststatistiken.

TAP-Kennwert	AUC	SE	p asymptotisch	CI (95%)	
				Untergrenze	Obergrenze
Alertness (A)					
Median RZ ohne Warnton (Ao Md)	.947	.030	< .0005***	.889	1.000
Standardabweichung ohne Warnton (Ao SD)	.967	.022	< .0005***	.923	1.000
Median RZ mit Warnton (Am Md)	.907	.038	< .0005***	.833	.982
Standardabweichung mit Warnton (Am SD)	.956	.028	< .0005***	.902	1.000
Kennwert phasische Alertness (Apha)	.747	.067	< .0005***	.616	.878
p (ES) Ao SD	.967	.022	< .0005***	.923	1.000
Geteilte Aufmerksamkeit (GA)					
auditiv Auslassungen (GAa A)	.930	.036	< .0005***	.861	1.000
auditiv Median RZ (GAa Md)	.945	.028	< .0005***	.889	1.000
auditiv Standardabweichung (GAa SD)	.875	.048	< .0005***	.780	.970
visuell Auslassungen (GAv A)	.819	.052	< .0005***	.717	.921
visuell Median RZ (GAv Md)	.900	.040	< .0005***	.822	.978
visuell Standardabweichung (GAv SD)	.939	.031	< .0005***	.877	1.000
Fehlreaktionen gesamt (GA F)	.805	.053	< .0005***	.702	.908
Auslassungen gesamt (GA A)	.954	.029	< .0005***	.897	1.000
p (ES) GA: aA, aMd, aSD, vMd, vSD, F, A	.997	.003	< .0005***	.990	1.000
p (ES) GA: aA, aMd, vSD	.993	.007	< .0005***	.979	1.000
Go/Nogo (GN)					
Median RZ (GN Md)	.816	.053	< .0005***	.713	.919
Standardabweichung (GN SD)	.950	.024	< .0005***	.904	.997
Fehlreaktionen (GN F)	.933	.038	< .0005***	.859	1.000
Auslassungen (GN A)	.860	.050	< .0005***	.761	.958
p (ES) GN: F, Md	.976	.017	< .0005***	.941	1.000
p (ES) GN: F, SD	.970	.020	< .0005***	.931	1.000

Anmerkungen: *** $p < .001$.

Die ROC-Kurven sämtlicher *TAP*-Kennwerte näherten sich der linken oberen Ecke an und visualisierten die hohe Diskriminationsfähigkeit der *TAP*-Kennwerte. Die größte Fläche unter der Kurve ($AUC = .997$) wurde mit dem Vorhersagemodell der Prädiktoren der *geteilten Aufmerksamkeit* mit Rückwärtsselektion ermittelt. Die *AUC*-Werte aller Vorhersagemodelle waren gemäß Hosmer und Lemeshow (2000) herausragend. Von den einzelnen *TAP*-Parametern ergab sich für den Kennwert *Ao SD* (*Alertness: Standardabweichungen der Reaktionszeiten ohne Warnton*) die größte Fläche unter der Kurve ($AUC = .967$). Die Kurvenverläufe zehn weiterer *TAP*-Parameter zeigten herausragende *AUC*-Werte. Die Trennschärfen von fünf *AUC*-Werten waren ausgezeichnet. Der Kennwert der *phasischen Alertness* (*Apha*) markierte mit einem *AUC*-Wert von $.747$ den flachsten Kurvenverlauf, der nach Hosmer und Lemeshow (2000) als akzeptabel zu bewerten war.

Aus den ROC-Analysen wurden die optimalen Schwellenwerte und Klassifikationsstatistiken ermittelt. Eine vollständige Auflistung der Schwellenwerte ist in Tabelle A-4 abgebildet. Tabelle 33 verdeutlicht die höchsten Klassifikationsstatistiken der *TAP*-Parameter für das Regressionsmodell mit Rückwärtsselektion diverser Kennwerte der *geteilten Aufmerksamkeit* (Schwellenwert $p = .436$) mit einem Youden's Index J von $.970$ und einer Klassifikationsgenauigkeit von 98.5% . Insgesamt wurden für die Schwellenwerte kombinierter Prädiktoren in den Regressionsmodellen eine bessere diagnostische Güte ermittelt als für einzelne *TAP*-Kennwerte. Eine Ausnahme stellte der *Alertness*-Kennwert *Standardabweichungen der Reaktionszeiten ohne Warnton* (*Ao SD*) mit einem Schwellenwert $T = 33.5$ dar. Dieser wies von sämtlichen einzelnen *TAP*-Kennwerten die beste Klassifikationsgenauigkeit von 95.5% (Sensitivität = 90.9% , Spezifität = 100%) auf und wurde als einziger Prädiktor in das Regressionsmodell der *Alertness* einbezogen, so dass gleiche Klassifikationsstatistiken resultierten. Abgesehen des Kennwerts *phasische Alertness* (*Apha*) rangierten die Sensitivitäts- und Spezifitätsmaße sämtlicher *Alertness*-Parameter oberhalb der BV-Konventionen. Von den Parametern der *geteilten Aufmerksamkeit* wurde mit dem Kennwert *auditive Auslassungen* (*Gaa A*) die zuverlässigste Gruppeneinteilung erzielt. Vergleichbare Klassifikationsgenauigkeiten von 92.4% wurden mit den *Alertness*-Parametern *Median der Reaktionszeiten ohne Warnton* (*Ao Md*) und *Standardabweichungen der Reaktionszeiten mit Warnton* (*Am SD*) ermittelt. Der Kennwert *Fehlreaktionen* (*GN F*) zeigte die höchste diagnostische Güte der *Inhibitionsfähigkeit* im *GoNogo*-Paradigma, 94% der Probanden wurden korrekt klassifiziert. Bei einer *PVV* von 100% ging die Chance, bei einer Testperformanz unterhalb des Schwellenwerts $T = 39.5$ negative Antwortverzerrungen zu

erfassen, gegen unendlich. Auch für die Kennwerte *Ao SD*, *Am SD*, *GAv Md*, *GA A* und *GN SD* lagen jeweils eine *PVV* von 100% vor. Die niedrigste *NVV* von 66.7% wies der Kennwert *visuelle Auslassungen der geteilten Aufmerksamkeit (GAv A)* auf, anhand dessen Schwellenwert von $T = 35.0$ knapp 50% der ES korrekt klassifiziert wurden, was einer Zufallseinteilung entsprach. Das niedrigste Quotenverhältnis *OR* von 9.0% wurde für den Kennwert *Fehlreaktionen der geteilten Aufmerksamkeit (GA F)* ermittelt.

Tabelle 33

Klassifikationsstatistiken der TAP-Kennwerte.

TAP-Kennwert	Cut-Off	Sensitivität	Spezifität	<i>PVV</i>	<i>NVV</i>	<i>HR</i>	<i>OR</i>	<i>J</i>
Alertness (A)								
Median RZ ohne Warnton (<i>Ao Md</i>)	31.5	90.9	93.9	93.7	91.2	92.4	153.8	84.8
Standardabweichung ohne Warnton (<i>Ao SD</i>)	33.5	90.9	100	100	91.7	95.5	Inf	90.9
Median RZ mit Warnton (<i>Am Md</i>)	31.5	78.8	97.9	96.3	82.1	87.9	120.2	75.8
Standardabweichung mit Warnton (<i>Am SD</i>)	35.5	84.8	100	100	86.8	92.4	Inf	84.8
Kennwert phasische Alertness (<i>Apha</i>)	50.5	66.7	87.9	84.6	72.5	77.3	14.6	54.6
<i>p</i> (ES) <i>Ao SD</i>	.720	90.9	100	100	91.7	95.5	Inf	90.9
Geteilte Aufmerksamkeit (GA)								
auditiv Auslassungen (<i>GAa A</i>)	34.0	84.8	97.0	96.6	86.5	90.9	180.4	81.8
auditiv Median RZ (<i>GAa Md</i>)	37.0	81.8	93.9	93.1	83.8	87.8	69.2	75.7
auditiv Standardabweichung (<i>GAa SD</i>)	41.0	87.9	97.0	96.7	88.9	92.4	234.9	84.9
visuell Auslassungen (<i>GAv A</i>)	43.0	75.8	100	100	80.5	87.9	Inf	75.8
visuell Median RZ (<i>GAv Md</i>)	46.5	87.9	90.9	90.6	88.3	89.4	72.6	78.8
visuell Standardabweichung (<i>GAv SD</i>)	35.0	51.5	97.0	94.5	66.7	74.2	34.3	48.5
Fehlreaktionen gesamt (<i>GA F</i>)	38.5	66.7	81.8	78.6	71.1	74.3	9.0	48.5
Auslassungen gesamt (<i>GA A</i>)	32.5	78.8	100	100	82.5	89.4	Inf	78.8
<i>p</i> (ES) <i>GA</i> : <i>aA</i> , <i>aMd</i> , <i>aSD</i> , <i>vMd</i> , <i>vSD</i> , <i>F</i> , <i>A</i>	.436	100	97.0	97.1	100	98.5	Inf	97.0
<i>p</i> (ES) <i>GA</i> : <i>aA</i> , <i>aMd</i> , <i>vSD</i>	.634	93.9	100	100	94.	96.9	Inf	93.9
Go/Nogo (GN)								
Median RZ (<i>GN Md</i>)	38.5	57.6	93.9	90.4	68.9	75.7	20.9	51.5
Standardabweichung (<i>GN SD</i>)	38.5	60.6	100	100	71.7	80.3	Inf	60.6
Fehlreaktionen (<i>GN F</i>)	39.5	87.9	100	100	89.2	94.0	Inf	87.9
Auslassungen (<i>GN A</i>)	38.5	75.8	97.0	96.2	80.0	86.4	101.3	72.8
<i>p</i> (ES) <i>GN</i> : <i>F</i> , <i>Md</i>	.579	93.9	93.9	93.9	93.9	93.9	237.0	87.8
<i>p</i> (ES) <i>GN</i> : <i>F</i> , <i>SD</i>	.594	90.9	97.0	96.8	91.4	94.0	323.0	87.9

7.5.4 Kombination der TAP- und CVLT-Kennwerte

Abschließend wurden die statistisch bedeutsamen *TAP*- und *CVLT*-Kennwerte der vorherigen Vorhersagemodelle in eine Regressionsanalyse mit Rückwärts- und Vorwärtsselektion eingeschlossen (*DG5*, *LS*, *FP*, *Ao SD*, *GAA A*, *GAA Md*, *GAA SD*, *GN F*, *GN Md*, *GN SD*).

Tabelle 34

TAP und CVLT: Binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion.

	Kennwert	<i>B</i>	<i>SE B</i>	Wald	<i>df</i>	<i>p</i>	<i>OR</i>	Klassifikation <i>n</i> (%)		
								total	ES	KG
rückwärts	GN SD	-1.52	317.31	.000	1	.996	.218	66 (100)	33 (100)	33 (100)
Schritt 9	LS	-1.54	269.32	.000	1	.995	.215			
	Konstante	133.96	18766.12	.000	1	.994	1.513E+58			
vorwärts	GN SD	-1.87	287.32	.000	1	.995	.154	66 (100)	33 (100)	33 (100)
Schritt 2	DG5	-2.94	339.12	.000	1	.993	.053			
	Konstante	187.10	20523.60	.000	1	.993	1.809E+81			

GN SD: GoNogo Standardabweichung, *CVLT*: Lernsumme *LS*, Durchgang 5 *DG5*.

Mit der logistischen Regressionsanalyse mit Rückwärtsselektion in Tabelle 34 konnten sämtliche Probanden korrekt klassifiziert werden. Nach Schritt 9 wurde das Vorhersagemodell mit den beiden Prädiktoren *Standardabweichungen der Reaktionszeiten* im *GoNogo*-Paradigma (*GN SD*) und *Lernsumme (LS)* des *CVLT* berechnet. Mit dem Vorhersagemodell konnte die Gesamtvarianz nach Schätzung gemäß Nagelkerke vollständig aufgeklärt werden ($\chi^2(2) = 91.495$, $p \leq .000$, Nagelkerke's $R^2 = 1.00$). Die korrigierte Gruppenvorhersage nach Firth lag weiterhin bei 100% für die beiden signifikanten Prädiktoren *LS* ($\chi^2 = 30.18$, $p \leq .000$) und *GN SD* ($\chi^2 = 6.38$, $p \leq .012$).

$$p(ES) \text{ rückwärts} = \frac{e [15.04 - .18(LS) - .17(GN SD)]}{1 + e [15.04 - .18(LS) - .17(GN SD)]}$$

Für das korrigierte Vorhersagemodell mit Rückwärtsselektion wurde eine sehr gute Interrater-Reliabilität mit dem *AKGT* und *WMT* von jeweils $\kappa = .970$ ($p < .000$) ermittelt.

Die logistische Regressionsanalyse mit Vorwärtsselektion nahm in Schritt 1 zunächst den Prädiktor *Durchgang 5 (DG5)* des *CVLT* und in Schritt 2 die *Standardabweichungen der Reaktionszeiten* im *GoNogo*-Paradigma (*GN SD*) in das Vorhersagemodell auf ($\chi^2(2) = 91.495$, $p \leq .000$, Nagelkerke's $R^2 = 1.00$). Dadurch konnte die Gruppenzugehörigkeit sämtlicher Probanden korrekt vorhergesagt werden. Mit dem korrigierten Vorhersagemodell

nach Firth konnten weiterhin alle Probanden durch die beiden signifikanten Prädiktoren *DG5* ($\chi^2 = 29.99, p \leq .000$) und *GN SD* ($\chi^2 = 10.26, p \leq .012$) korrekt klassifiziert werden.

$$p(ES)_{\text{vorwärts}} = \frac{e [15.68 - .22(DG5) - .17(GN SD)]}{1 + e [15.68 - .22(DG5) - .17(GN SD)]}$$

Die Übereinstimmung des korrigierten Vorhersagemodells mit dem *WMT* und dem *AKGT* betrug jeweils $\kappa = .970$ ($p < .000$).

Im nächsten Schritt wurden die optimalen Schwellenwerte und die Klassifikationsstatistiken der Wahrscheinlichkeiten für beide Regressionsmodelle p (ES) berechnet. Die ROC-Analysen ergaben für beide Vorhersagemodelle jeweils eine perfekte AUC von 1.0 (CI: 1.00 – 1.00, $p < .000$). Für das Regressionsmodell mit Rückwärtsselektion und den beiden Prädiktoren *LS* und *GN SD* wurden bei einem Schwellenwert von $p = .548$ eine Sensitivität und Spezifität von 100% ermittelt. Bei einem Schwellenwert von $p = .448$ ergab sich auch für das Regressionsmodell mit Vorwärtsselektion (*DG5* und *GN SD*) eine perfekte Klassifikationsgenauigkeit von 100%.

Zusammenfassend wurden die Kennwerte der Modellgüte der einzelnen Vorhersagemodelle in Tabelle 35 aufgeführt und um die beiden Kriterien AIC und BIC für die Modellanpassung ergänzt.

Tabelle 35

Kennwerte der Modellgüte und Modellanpassung der logistischen Regressionsmodelle.

	Prädiktoren	R^2_N	Kappa _{WMT}	Kappa _{AKGT}	AIC	BIC	AUC
CVLT: p (ES) rückwärts	LS, FP	1.00	.939	.939	66.00	138.26	.999
CVLT: p (ES) vorwärts	DG5, FP	1.00	.939	.939	48.00	100.55	.999
TAP: p (ES) rück-/ vorwärts	Ao SD	.84	.879	.818	69.55	139.61	.967
TAP: p (ES) rückwärts	GAa A, GAa Md, GAa SD, GAv Md, GAv SD; GA F, GA A	1.00	.909	.909	132.00	276.52	.997
TAP: p (ES) vorwärts	GAa A, GAa Md, GAv SD	.93	.848	.848	128.00	268.14	.993
TAP: p (ES) rückwärts	GN F, GN Md	.85	.879	.818	97.39	202.49	.976
TAP: p (ES) vorwärts	GN F, GN SD	.84	.879	.818	91.39	189.92	.970
TAP/CVLT: p (ES) rückwärts	GN SD, LS	1.00	.970	.970	112.00	234.62	1.00
TAP/CVLT: p (ES) vorwärts	GN SD, DG5	1.00	.970	.970	94.00	196.91	1.00

Anmerkungen: Legende siehe Seite 149.

Tabelle 35 verdeutlicht, dass mit sämtlichen Modellen zuverlässige Gruppenvorhersagen ermittelt werden konnten. Dabei ergaben sich die höchsten Kennwerte für die beiden Modelle mit den kombinierten Prädiktoren aus *TAP* und *CVLT*.

7.5.5 Ergebnisse zu Fragestellung F4 und Hypothese H6

F4: Inwiefern eignen sich verbal-mnestische und attentionale Testvariablen neuropsychologischer Standardverfahren zur Vorhersage der Gruppenzugehörigkeit?

H6: Die experimentelle Gruppenzugehörigkeit kann zuverlässig aus dem *TAP*-Untertest *Alertness* vorhergesagt werden. Auch die Parameter der *TAP*-Untertests der *geteilten Aufmerksamkeit* und *Inhibitionsfähigkeit* können zur Gruppenvorhersage herangezogen werden.

Die Ergebnisse bestätigen die Annahme, dass sich die Gruppenzugehörigkeit zuverlässig mit *TAP*-Kennwerten vorhersagen lässt. Eine zuverlässige Vorhersage mit Kennwerten der *geteilten Aufmerksamkeit* war lediglich durch eine Kombination diverser Kennwerte möglich. Dagegen konnten 92.4% der Probanden mit dem *Alertness*-Kennwert *Standardabweichung der Reaktionszeiten ohne Warnton (Ao SD)* vorhergesagt werden. Die Kombination des *GoNogo*-Parameters *Fehlreaktionen (GN F)* mit jeweils dem *Median der Reaktionszeiten (GN Md)* sowie *Fehlreaktionen (GN F)* und *Standardabweichung der Reaktionszeiten (GN SD)* ermöglichte ebenfalls eine korrekte Vorhersage von 92.4%. Während die *Inhibitionsfähigkeit* zuverlässig und effizient zur Gruppenvorhersage beitrug, konnte eine zuverlässige Vorhersage nur mit einer Kombination aus diversen Kennwerten der *geteilten Aufmerksamkeit* erzielt werden. Insofern kann H6 nur teilweise als bestätigt gelten.

Die besten Klassifikationsstatistiken wurden für die kombinierten Vorhersagemodelle mit *TAP*- und *CVLT*-Kennwerten ermittelt. Die Regressionsmodelle mit den Prädiktoren *Standardabweichung der Reaktionszeiten GoNogo (GN SD)* und jeweils einem der beiden *CVLT*-Kennwerte *Durchgang 5 (DG5)* und *Lernsumme (LS)* wiesen jeweils eine perfekte Gruppenvorhersage auf.

8 Ergebnisse: zweiter Teil der Studie

Der Einbezug einer psychiatrisch-neurologischen Patientenstichprobe mit negativen Antwortverzerrungen (PNA) ermöglichte den Vergleich der mittleren Testperformanz mit den ES und eine Validierung der ermittelten Schwellenwerte.

8.1 Deskriptive Ergebnisdarstellung

Zur Veranschaulichung wurde die mittlere Testperformanz der PNA und ES graphisch dargestellt (Abbildung 7, Abbildung 8).

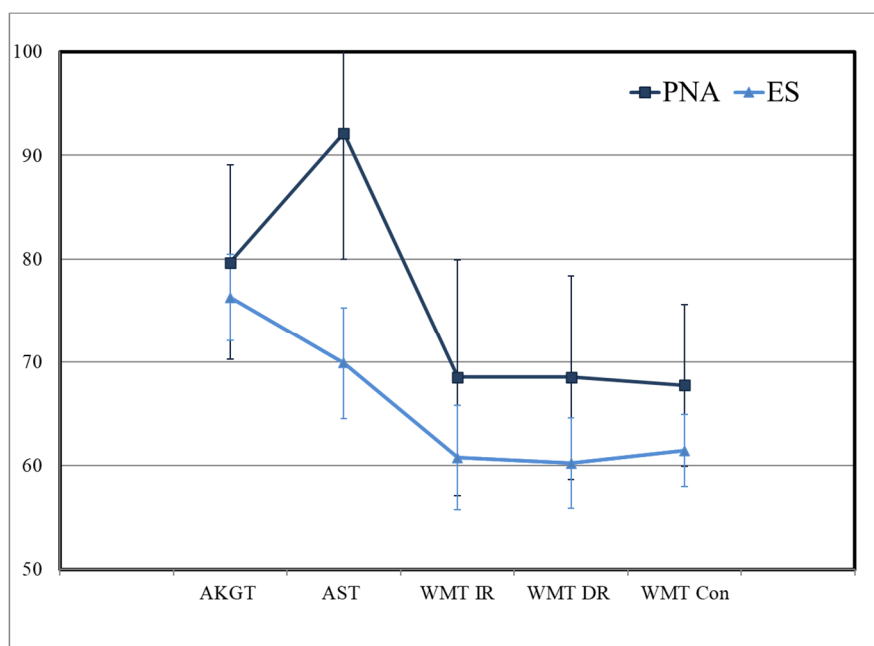


Abbildung 7. Mittlere Prozentwerte mit 95%-Konfidenzintervallen der PNA ($n = 10$) und ES ($n = 33$) im Amsterdamer Kurzzeitgedächtnistest (AKGT), Aggravations- und Simulationstests (AST), Word Memory Test (WMT): unmittelbares Wiedererkennen (WMT IR), verzögertes Wiedererkennen (WMT DR) und Konsistenzmaß (WMT Con).

Abbildung 7 verdeutlicht, dass die PNA in sämtlichen PVT eine bessere mittlere Testperformanz als die ES zeigten. Mit Ausnahme des *Aggravations- und Simulationstests* (AST) überschritten sich die 95%-Konfidenzintervalle der ES und PNA in allen anderen PVT, so dass lediglich im AST (PNA: $M = 92.14$; ES: $M = 69.91$) von einem statistisch bedeutsamen Unterschied auszugehen ist. Während die mittlere Testperformanz der ES im AST deutlich unterhalb des kritischen Schwellenwerts von 90% rangierte, lag der mittlere AST-Wert der PNA oberhalb des Schwellenwerts.

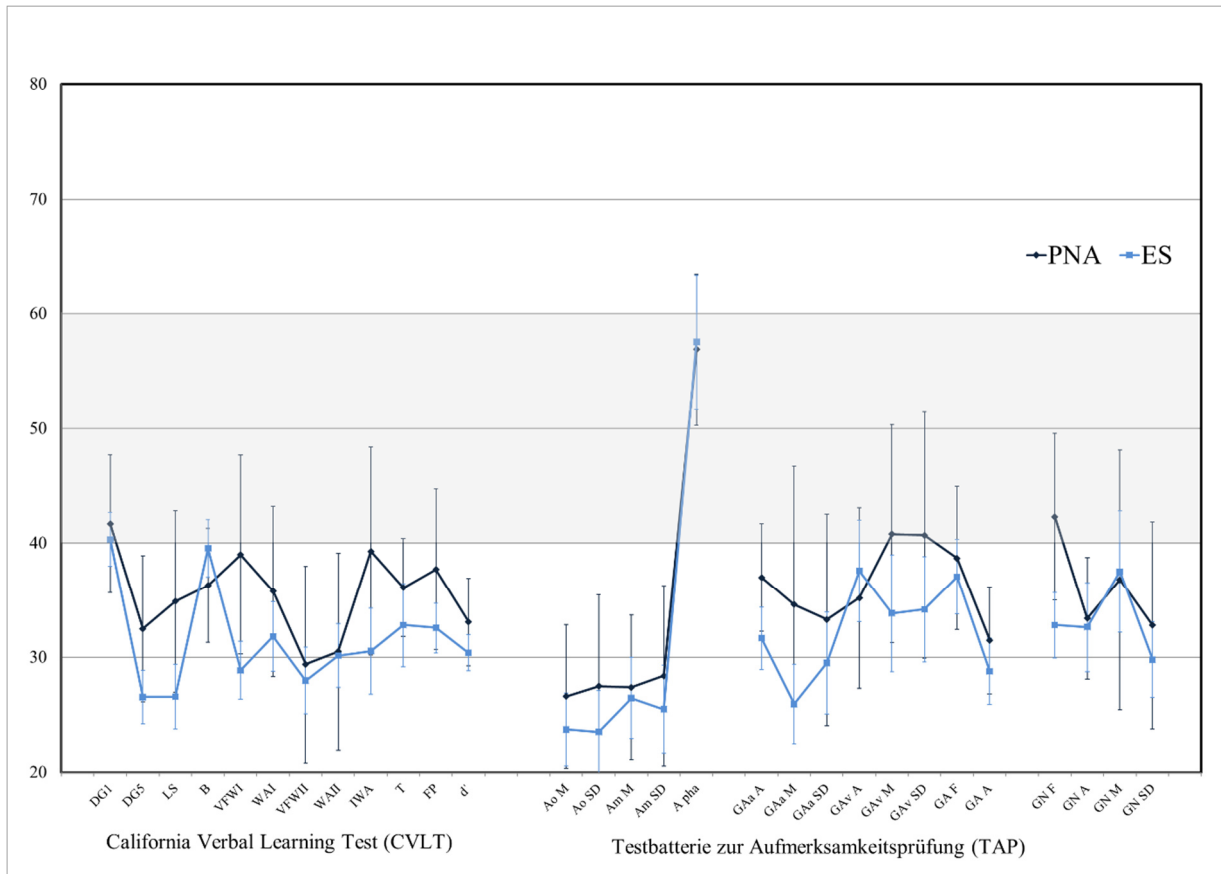


Abbildung 8. Mittlere T-Werte mit 95%-Konfidenzintervallen der PNA ($n = 10$) und ES ($n = 33$): California Verbal Learning Test (CVLT) und Testbatterie zur Aufmerksamkeitsprüfung (TAP). Graue Fläche markiert Normbereich ($40 \leq T \leq 60$).

CVLT-Untertests: *DG1*: Liste A Durchgang 1, *DG5*: Liste A Durchgang 5, *LS*: Liste A Lernsumme, *B*: Interferenzliste B, *VFWI*: Verzögerte freie Wiedergabe I, *WAI*: Wiedergabe mit Abrufhilfe I, *VFWII*: Verzögerte freie Wiedergabe II, *WAI*: Wiedergabe mit Abrufhilfe II, *IWA*: Intrusionen während Wiedergabe mit Abrufhilfe, *T*: Trefferrate bei Wiedererkennen von Liste A, *FP*: Falsch Positive bei Wiedererkennen von Liste A, *d'*: Diskriminationswert bei Wiedererkennen Liste A.

TAP-Untertests:

Alertness: *AoM*: Median RZ ohne Warnton, *AoSD*: Standardabweichung ohne Warnton, *AmM*: Median RZ mit Warnton, *AmSD*: Standardabweichung mit Warnton, *Apha*: Kennwert phasische Alertness.

Geteilte Aufmerksamkeit: *GAA*: auditiv Auslassungen, *GAA*: auditiv Median RZ, *GAA*: auditiv Standardabweichung, *GAA*: visuell Auslassungen, *GAA*: visuell Median RZ, *GAA*: visuell Standardabweichung, *GAF*: Fehlreaktionen gesamt, *GAA*: Auslassungen gesamt.

GoNogo: *GNF*: Fehlreaktionen, *GNA*: Auslassungen, *GNA*: Median RZ, *GNA*: Standardabweichung.

Aus Abbildung 8 geht hervor, dass die Testperformanz beider Stichproben im CVLT und in der TAP mehrheitlich im unterdurchschnittlichen Bereich lag. Die markanteste Ausnahme stellte der Kennwert der phasischen Alertness (*Apha*) dar, der für beide Gruppen im oberen Normbereich lag (PNA: $M = 56.90$; ES: $M = 57.55$). Im Gegensatz zu den Parametern der ES ragten einzelne Kennwerte der PNA in den Durchschnittsbereich: *Durchgang 1 (DG1*:

$M = 41.70$) des *CVLT*, *Median der Reaktionszeiten visuell* (*GA_v Md*: $M = 40.80$) und *Standardabweichungen der Reaktionszeiten visuell* (*GA_v SD*: $M = 40.70$) der *geteilten Aufmerksamkeit* sowie *Fehlreaktionen GoNogo* (*GN F*: $M = 42.30$). Darüber hinaus zeigten die ES in zahlreichen Parametern eine schlechtere Testperformanz als die PNA. Allerdings überschritten sich die Konfidenzintervalle in sämtlichen Kennwerten, im *GoNogo*-Parameter *Fehlreaktionen* (*GN F*) nur knapp. Lediglich in vier Kennwerten zeigten die ES eine bessere bzw. vergleichbare Testperformanz als die PNA: *Interferenzliste B* des *CVLT* (PNA: $M = 36.30$; ES: $M = 39.55$), *physische Alertness* (*Apha*: PNA: $M = 56.90$; ES: $M = 57.55$), *Auslassungen visuell* der *geteilten Aufmerksamkeit* (*GA_v A*: PNA: $M = 35.20$; ES: $M = 37.58$) und *Median der Reaktionszeiten GoNogo* (*GN Md*: PNA: $M = 36.80$; ES: $M = 36.97$).

8.2 Performanzvaliditätstests

8.2.1 Inferenzstatistische Gruppenunterschiede

Tabelle 36

Deskriptive Statistiken und Effektstärkenmaße der PVT-Gruppenvergleiche.

PVT %	PNA ($n = 10$)		ES ($n = 33$)		t	df	p	<i>Cohens d</i>
	M	SD	M	SD				
AKGT	79.67	14.53	76.26	11.76	-.759	41	.452	-0.27
AST	92.14	10.74	69.91	15.13	-4.311	41	.000**	-1.56
WMT IR	68.50	16.96	60.76	14.20	-1.444	41	.156	-0.52
WMT DR	68.50	21.09	60.23	12.27	-1.563	41	.126	-0.56
WMT Con	67.75	17.14	61.44	9.78	-1.111	11	.291	-0.40

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur ** $p < .010$.

Aus den Mittelwertvergleichen in Tabelle 36 geht hervor, dass sich die PNA und ES hoch signifikant im *AST* voneinander unterschieden. Die PNA präsentierten dabei weniger Fehler als die ES. Im *AKGT* und *WMT* wurden keine statistisch bedeutsamen Differenzen ermittelt.

8.2.2 Klassifikation mit dem AST

Tabelle 37

Klassifikationen der ES und PNA mit AST-Schwellenwerten.

	PNA ($n = 10$)		ES ($n = 33$)	
	Richtig Positiv n	Falsch Negativ n	Richtig Positiv n	Falsch Negativ n
AST \leq 90.0%	3 (30.0%)	7 (70.0%)	30 (90.9%)	3 (9.1%)
AST \leq 95.7%	4 (40.0%)	6 (60.0%)	32 (97.0%)	1 (3.0%)

Wie Tabelle 37 verdeutlicht, konnten 90.9% der ES mit einem Schwellenwert von 90% des *AST* korrekt klassifiziert werden. Dagegen lag der Anteil Richtig-Positiver-Entscheidungen der PNA bei lediglich 30%. Die Beobachterübereinstimmung zwischen dem *AST* \leq 90.0% und dem *WMT* und *AKGT* unterschied sich jeweils nicht signifikant von einer zufallsbedingten Übereinstimmung (je $\kappa = .359$, $p < .070$). Durch das Erhöhen des *AST* Schwellenwerts auf 95.7% konnte in der klinischen Stichprobe die Anzahl Richtig-Positiver-Entscheidungen auf 40% verbessert werden. Daraus ergab sich jeweils eine mäßig bis gute Beobachterübereinstimmung von $\kappa = .609$ ($p < .010$) für den *AST* mit dem *WMT* und *AKGT*.

8.2.3 Ergebnisse zu Fragestellung F5 und Hypothese H7

F5: Bestehen Unterschiede in der Testperformanz zwischen experimentellen Simulanten (ES) und psychiatrisch-neurologischen Patienten mit negativen Antwortverzerrungen (PNA)?

H7: Es bestehen statistisch bedeutsame Unterschiede im *AST* zwischen den experimentellen Simulanten (ES) und der klinischen PNA-Stichprobe. Mit dem *AST* können weniger Richtig-Positiv-Entscheidungen der PNA getroffen werden.

Im Vergleich zu den ES zeigten die PNA eine bessere Testperformanz im *AST*, während keine statistisch bedeutsamen Gruppenunterschiede im *AKGT* und *WMT* ermittelt wurden. Darüber hinaus wurden mit dem *AST* hypothesenkonform eine geringere Anzahl an Richtig-Positiv-Entscheidungen im klinischen als im experimentellen Kontext ermittelt. Während im experimentellen Kontext 90.0% (Schwellenwert 90%) bzw. 97.0% (Schwellenwert 95.7%) der ES korrekt klassifiziert wurden, lag die Rate korrekt klassifizierter negativer Antwortverzerrungen der PNA bei 30.0% (Schwellenwert 90%) bzw. 40.0% (Schwellenwert 95.7%). Insgesamt kann Hypothese H7 als bestätigt gelten.

8.3 Neuropsychologische Standardverfahren

8.3.1 Inferenzstatistische Gruppenunterschiede

Tabelle 38

Deskriptive Statistiken und Effektstärken der Gruppenunterschiede im CVLT.

CVLT	PNA (<i>n</i> = 10)		ES (<i>n</i> = 33)		<i>t</i>	<i>df</i>	<i>p</i>	Cohens <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Durchgang 1 (DG1)	41.70	8.41	40.33	6.62	-.537	41	.594	-0.19
Durchgang 5 (DG 5)	32.50	8.99	26.55	6.59	-2.305	41	.026	-0.57
Lernsumme (LS)	36.30	6.98	39.55	7.08	1.274	41	.210	0.46
Interferenzliste B (B)	34.90	11.12	26.58	7.96	-2.636	41	.012	-0.65
Verzögerte freie Wiedergabe I (VFW I)	39.00	12.12	28.88	7.15	-2.512	11	.029	-1.19
Wiedergabe mit Abrufhilfe I (WA I)	35.80	10.40	31.82	8.53	-1.229	41	.226	-0.44
Verzögerte freie Wiedergabe II (VFW II)	29.40	11.97	27.97	8.17	-.433	41	.667	-0.11
Wiedergabe mit Abrufhilfe II (WA II)	30.50	12.03	30.15	7.81	-.108	41	.914	-0.03
Intrusionen Wiedergabe Abrufhilfe (IWA)	39.30	12.72	35.88	10.62	-.853	41	.399	-0.31
Trefferrate (T)	36.10	6.027	32.82	10.22	-.961	41	.342	-0.24
Falsch Positive (FP)	37.70	9.83	32.58	6.13	-1.996	11	.053	-0.49
Diskriminationswert (<i>d'</i>)	33.10	5.34	30.39	4.43	-1.614	41	.114	-0.40

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$.

Die größten Mittelwertdifferenzen im CVLT zeigten sich nach Tabelle 38 in den Parametern *Lernsumme (LS)*, *Durchgang 5 (DG5)* und (kurz) *verzögerte freie Wiedergabe I (VFW I)*. Die PNA präsentierten bessere Testleistungen als die ES. In der Signifikanzprüfung nach Bonferroni-Holm-Korrektur des α -Fehlers erwiesen sich die Leistungsunterschiede jedoch nicht mehr als statistisch bedeutsam.

Aus den Mittelwertvergleichen der TAP-Kennwerte in Tabelle 39 geht hervor, dass die beiden Gruppen keine statistisch bedeutsamen Leistungsdifferenzen in den Parametern der *Alertness* und der *geteilten Aufmerksamkeit* zeigten. Lediglich in der Anzahl an *Fehlreaktionen* im *GoNogo-Paradigma* der *Inhibitionsfähigkeit (GN F)* unterschieden sich die beiden Gruppen. Die ES produzierten signifikant mehr Fehlreaktionen als die PNA ($d = -1.11$).

Tabelle 39

Deskriptive Statistiken und Effektstärken der Gruppenunterschiede in den TAP-Untertests.

TAP	PNA (<i>n</i> = 10)		ES (<i>n</i> = 33)		<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Alertness (A)								
Median RZ ohne Warnton (Ao Md)	26.60	8.72	23.73	8.92	-0.896	41	.375	-0.32
Standardabweichung ohne Warnton (Ao SD)	27.50	11.16	23.52	10.07	-1.070	41	.291	-0.39
Median RZ mit Warnton (Am Md)	27.40	8.81	26.45	9.97	-0.269	41	.789	-0.10
Standardabweichung mit Warnton (Am SD)	28.40	11.00	25.48	10.79	-0.746	41	.460	-0.27
Kennwert phasische Alertness (Apha)	56.90	9.21	57.55	16.54	.117	28	.907	0.04
Geteilte Aufmerksamkeit (GA)								
auditiv Auslassungen (GAa A)	34.60	16.93	25.94	9.73	-1.543	11	.152	-0.56
auditiv Median RZ (GAa Md)	33.33	12.91	29.52	12.57	-0.829	41	.412	-0.30
auditiv Standardabweichung (GAa SD)	37.00	6.58	31.67	7.67	-1.985	41	.054	-0.72
visuell Auslassungen (GAv A)	40.80	13.32	33.39	14.19	-1.465	41	.151	-0.53
visuell Median RZ (GAv Md)	40.70	15.05	33.73	12.83	-1.444	41	.156	-0.52
visuell Standardabweichung (GAv SD)	35.20	11.06	37.58	12.87	.540	41	.592	0.20
Fehlreaktionen gesamt (GA F)	38.70	8.72	37.06	9.16	-0.501	41	.619	-0.18
Auslassungen gesamt (GA A)	31.50	6.52	28.79	8.16	-0.960	41	.343	-0.35
Go/Nogo (GN)								
Median RZ (GN Md)	36.80	15.84	36.97	14.86	.031	41	.975	-0.01
Standardabweichung (GN SD)	32.80	12.64	29.48	9.22	-0.912	41	.367	-0.33
Fehlreaktionen (GN F)	42.30	10.19	32.82	8.07	-3.062	41	.004*	-1.11
Auslassungen (GN A)	33.40	7.41	32.64	10.92	-0.206	41	.838	-0.07

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur * $p < .05$.

8.3.2 Klassifikationen mit dem CVLT und der TAP

Zur Vorhersage der Gruppenzugehörigkeit wurden neben den (korrigierten) logistischen Regressionsmodellen und deren optimalen Schwellenwerten aus den ROC-Analysen auch die Schwellenwerte einzelner und kombinierter Kennwerte herangezogen. In Tabelle 40 werden die *CVLT* und *TAP*-Kennwerte mit der höchsten Anzahl Richtig-Positiver-Klassifikationen aufgeführt, eine vollständige Auflistung sämtlicher Einteilungen findet sich in Tabelle A-5 im Anhang.

Tabelle 40

Klassifikationen der ES und PNA mit CVLT und TAP-Kennwerten.

	PNA ($n = 10$)		ES ($n = 33$)	
	Richtig Positiv	Falsch Negativ	Richtig Positiv	Falsch Negativ
	n (%)	n (%)	n (%)	n (%)
CVLT und TAP				
p (ES) LS, GN SD $\geq .500$	9 (90.0)	1 (10.0)	33 (100)	0 (0)
p (ES) LS, GN SD $\geq .548$	9 (90.0)	1 (10.0)	33 (100)	0 (0)
p (ES) DG5, GN SD $\geq .500$	9 (90.0)	1 (10.0)	33 (100)	0 (0)
p (ES) DG5, GN SD $\geq .448$	9 (90.0)	1 (10.0)	33 (100)	0 (0)
LS ≤ 42.5 und GN SD ≤ 38.5	10 (100)	0 (0)	33 (100)	0 (0)
DG5 ≤ 36.0 und GN SD ≤ 38.5	10 (100)	0 (0)	33 (100)	0 (0)

CVLT: DG5: Durchgang 5, LS: Lernsumme, GoNogo: GN SD: Standardabweichung.

Tabelle 40 verdeutlicht, dass mit den kombinierten Prädiktoren *Standardabweichungen der Reaktionszeiten* im GoNogo-Paradigma (GN SD) und *Lernsumme (LS)* oder *Durchgang 5 (DG5)* des CVLT die höchste Anzahl Richtig-Positiver-Entscheidungen ermittelt werden konnte. Dies zeigte sich nicht nur in der experimentellen, sondern auch in der klinischen Stichprobe. Dabei erwies sich in der PNA-Stichprobe die Verwendung kombinierter Schwellenwerte $LS \leq 42.5$ und $GN SD \leq 38.5$ bzw. $DG5 \leq 36.0$ und $GN SD \leq 38.5$ als zuverlässigeres Vorgehen im Vergleich zu den entsprechenden logistischen Vorhersagemodellen, da die Rate Richtig-Positiver-Entscheidungen von 90% auf 100% verbessert werden konnte.

8.3.3 Ergebnisse zu Fragestellung F5 und Hypothese H8

F5: Bestehen Unterschiede in der Testperformanz zwischen experimentellen Simulanten (ES) und psychiatrisch-neurologischen Patienten mit negativen Antwortverzerrungen (PNA)?

H8: Es bestehen statistisch bedeutsame Unterschiede im CVLT und der TAP zwischen den experimentellen Simulanten (ES) und der klinischen PNA-Stichprobe. Mit den eingebetteten Parametern können weniger Richtig-Positiv-Entscheidungen der PNA getroffen werden.

Mit Ausnahme einer höheren Anfälligkeit für Fehlreaktionen der ES im GoNogo-Paradigma der *Inhibitionsfähigkeit* ($d = -1.11$) zeigten sich entgegen Hypothese H8 keine statistisch bedeutsamen Unterschiede in der Testperformanz der ES und PNA in den verbalen Gedächtnisleistungen und Aufmerksamkeitsfunktionen. Die Gruppenzugehörigkeit der PNA konnte am besten mit den beiden Prädiktorkombinationen aus CVLT und TAP

Standardabweichungen der Reaktionszeiten im GoNogo-Paradigma (GN SD) und Lernsumme (LS) sowie Standardabweichungen der Reaktionszeiten im GoNogo-Paradigma (GN SD) und Durchgang 5 (DG5) vorhergesagt werden. Die höchste Anzahl an Richtig-Positiven-Entscheidungen von 100% wurde jeweils mit einer Kombination der beiden Schwellenwerte ermittelt.

9 Ergebnisse: dritter Teil der Studie

Im dritten Teil der vorliegenden Arbeit wurden die Grenzen des *AKGT* und *AST* untersucht. Dazu wurde die Testperformanz psychiatrisch-neurologischer Patienten mit negativen Antwortverzerrungen (PNA) und Patienten mit tatsächlichen kognitiven Beeinträchtigungen (PKB) verglichen. Ferner wurden die Zusammenhänge zwischen den PVT und den Untertests der *CERADplus-NTB* untersucht, um die Funktionsbereiche zu ermitteln, die während der Bearbeitung der beiden PVT beansprucht werden.

9.1 Performanzvalidierungstests

Anhand multipler *t*-Tests für unabhängige Stichproben mit Bonferroni-Holm-Korrektur wurden die Gruppenunterschiede in der ersten Testhälfte des *Amsterdamer Kurzzeitgedächtnistests (AKGT 1)* und des *Aggravations- und Simulationstests (AST)* ermittelt.

Tabelle 41

Deskriptive Statistiken und Effektstärken der PVT-Gruppenvergleiche.

	PNA (<i>n</i> = 10)		PKB (<i>n</i> = 12)		<i>t</i>	<i>df</i>	<i>p</i>	<i>Cohens d</i>
	<i>MW</i>	<i>SD</i>	<i>MW</i>	<i>SD</i>				
AKGT 1	35.90	6.28	36.75	4.07	-.383	20	.706	0.16
AST	92.14	10.74	85.36	12.08	1.379	20	.183	-0.59
AST Md	1381.40	237.36	1980.42	1066.12	-1.891	12	.082	0.78
AST SD	1706.20	844.08	5029.42	3951.21	-2.837	12	.015	1.16

Anmerkungen. Signifikant nach Bonferroni-Holm-Korrektur.

Aus Tabelle 41 wird ersichtlich, dass zwischen den Gruppen keine statistisch bedeutsamen Mittelwertsdifferenzen vorlagen. Zwar erzielten die PKB schlechtere Erinnerungsleistungen (*AGKT1*, *AST*) bei längeren Reaktionszeiten (*AST Md*) und größeren Standardabweichungen (*AST SD*) als die PNA, jedoch zeigten Leistungsdifferenzen nach der Korrektur gemäß Bonferroni-Holm keine statistische Signifikanz.

Anschließend wurden die PKB und PNA anhand der Schwellenwerte des *AKGT 1* und des *AST* in die beiden Gruppen „bestanden“ und „nicht bestanden“ klassifiziert. Neben den Schwellenwerten der jeweiligen Testhandbücher wurden auch die optimalen Schwellenwerte der vorliegenden Analogstudie herangezogen und die Klassifikationsstatistiken unter Berücksichtigung einer Basisrate negativer Antwortverzerrungen von $p = .45$ ermittelt.

Tabelle 42

Klassifikationsstatistiken der *AKGT*- und *AST*-Parameter.

PVT	Sensitivität	Spezifität	<i>PVV</i>	<i>NVV</i>	<i>HR</i>	<i>OR</i>
<i>AKGT</i> $1 \leq 39$	70.0	25.0	43.0	49.4	44.7	0.8
<i>AST</i> $\leq 90.0\%$	30.0	41.7	29.6	42.1	36.4	0.3
<i>AST</i> $\leq 95.7\%$	50.0	16.7	32.9	29.0	31.7	0.2

Tabelle 42 verdeutlicht, dass die Gruppeneinteilung der PKB mit den Schwellenwerten des *AKGT1* und *AST* zu zahlreichen Fehlklassifikationen führte. Im *AKGT 1* erzielten 70% der PNA und 75% der PKB eine auffällige Testperformanz, so dass lediglich 25% der PKB als Richtig-Negative klassifiziert wurden. Die Klassifikationsgenauigkeit von 44.7% lag unterhalb der Zufallswahrscheinlichkeit. Auch mit den beiden Schwellenwerten 90% und 95.7% des *AST* wurden Klassifikationsgenauigkeiten unterhalb der Ratewahrscheinlichkeit (36.4% und 31.7%) ermittelt. Für den *AST* (Schwellenwert $\leq 95.7\%$) lag neben dem niedrigsten Quotenverhältnis *OR* von 0.2% auch die niedrigste *NVV* von 29.0% vor.

9.1.1 Ergebnisse zu Fragestellung F6 und Hypothese H9

F6: Bestehen Unterschiede im *AKGT* und *AST* zwischen den Patienten mit verminderter Anstrengungsbereitschaft (PNA) und den Patienten mit authentischen kognitiven Beeinträchtigungen (PKB)?

H9: Es zeigen sich keine statistisch bedeutsamen Unterschiede zwischen den beiden Patientenstichproben im *AKGT* und *AST*. Bei den Patienten mit authentischen kognitiven Beeinträchtigungen (PKB) ist von einer erhöhten Anzahl an Falsch-Positiv-Fehlklassifikationen auszugehen.

In der vorliegenden Arbeit konnten die Leistungseinbußen tatsächlicher kognitiver Dysfunktionen und negativer Antwortverzerrungen im *AST* und *AKGT 1* nicht voneinander unterschieden werden. Daraus resultierten zahlreiche Falsch-Positive-Fehlklassifikationen durch den *AKGT1* (75%) und *AST* (58.3% bzw. 83.3%). Demzufolge wurden die Annahmen von Hypothese H9 erfüllt.

9.2 Kognitive Funktionen und PVT

In einem weiteren Schritt wurden die kognitiven Funktionen der PKB deskriptiv in Tabelle 43 dargestellt. Die Zusammenhänge zwischen den Parametern der *CERADplus-NTB* und des *AKGT 1* und *AST* wurden anhand von Korrelationen nach Pearson (Tabelle 44) verdeutlicht.

Tabelle 43

Deskriptive Statistiken: z-standardisierte Parameter der CERADplus-NTB der PKB (n = 12).

	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min.</i>	<i>Max</i>
VF	-2.01	1.14	-1.73	-5.10	-.80
BNT	-1.11	2.33	-.94	-6.70	1.60
MMSE	-2.85	2.06	-2.55	-7.80	.40
WLG	-2.56	1.50	-2.50	-6.00	-.90
DG1	-1.74	.91	-1.51	-4.00	-.60
DG2	-1.91	1.44	-1.65	-4.70	.19
DG3	-2.60	1.57	-2.49	-5.90	-.50
WLA	-2.38	1.11	-2.10	-4.80	-1.26
WLB	-2.20	1.23	-1.65	-4.90	-.69
WLI	-.68	1.21	-.75	-2.70	.81
D	-1.49	1.60	-1.85	-3.90	.95
FAZ	-2.31	2.29	-1.81	-8.40	.30
FA	-1.96	1.27	-2.15	-4.10	-.10
FB	-.83	1.80	-1.60	-3.21	3.00
TMT A	-1.70	1.12	-2.00	-3.20	.10
PF	-.68	1.23	-.65	-3.50	1.10

Anmerkungen: CERAD-Plus: VF: Verbale Wortflüssigkeit, BNT: Boston Naming Test, MMSE: Mini Mental Status Examination, WLG: Wortliste Gedächtnis gesamt, DGI-3: Wortlistendurchgänge 1, 2 und 3, WLA: Wortliste abrufen, WLB: Wortliste Behalten, WLI: Wortliste Intrusionen, D: Diskriminabilität Wortliste Wiedererkennen, FAZ: Figuren Abzeichnen, FA: Figuren Abrufen, FB: Figuren Behalten, TMT A: Trail Making Test Teil A, PF: Phonematische Wortflüssigkeit.

Tabelle 44.

Korrelationskoeffizienten *n*. Spearman der PKB (*n* = 12): AKGT, AST und CERADplus-NTB.

	VF	BNT	MMSE	WLG	DG1	DC2	DG3	WLA	WLI	WLB	D	FAZ	FA	FB	PF	TMTA	AKGT1	AST
VF	1																	
BNT	.340	1																
MMSE	.508	.256	1															
WLG	.699*	.490	.637*	1														
DG1	.633*	.637*	.785**	.844**	1													
DC2	.466	.291	.640*	.893**	.660*	1												
DG3	.807**	.271	.633*	.909**	.781**	.747**	1											
WLA	.640*	.099	.535	.847**	.542	.914**	.820**	1										
WLI	.274	.135	.044	.070	.224	-.253	.162	-.272	1									
WLB	.435	.225	.332	.684*	.478	.798**	.623*	.797**	-.312	1								
D	.028	.550	.102	.427	.302	.529	.225	.380	-.358	.639*	1							
FAZ	.000	-.137	.249	.091	.274	-.049	.147	-.137	.481	-.330	-.552	1						
FA	.608*	.445	.627*	.671*	.647*	.616*	.596*	.552	-.049	.386	.266	.273	1					
FB	.424	.288	.456	.424	.335	.523	.359	.509	-.355	.585*	.602*	-.459	.592*	1				
PF	.636*	.371	.462	.657*	.524	.571	.733**	.615*	-.218	.474	.322	-.056	.727**	.525	1			
TMTA	.746**	.134	.642*	.570	.556	.399	.693*	.435	.484	.148	-.190	.465	.711**	.250	.0563	1		
AKGT1	.541	.254	.437	.566	.360	.570	.541	.548	.104	.319	.193	.137	.721**	.298	.633*	.758**	1	
AST	.604*	-.018	.460	.611*	.496	.402	.718**	.459	.474	.072	-.358	.642*	.512	-.076	.467	.834**	.545	1

r < .10 *r* < .30 *r* < .50 *r* < .70 *r* > .70

Anmerkung: * *p* < .050 und ** *p* < .010.

Die Rangkorrelationen nach Spearman in Tabelle 44 verdeutlichen jeweils einen statistisch hoch bedeutsamen Zusammenhang zwischen den Leistungen im *AKGT 1* und *AST* und der zentralnervösen Informationsverarbeitungsgeschwindigkeit (*TMT A: Trail Making Test A*). Darüber hinaus zeigte sich ein hoch signifikanter Zusammenhang zwischen dem *AKGT 1* und den figuralen Abrufleistungen (*FA: Figuren Abrufen*). Ferner korrelierte der *AKGT 1* mit den exekutiven Funktionsbereichen, im Sinne der *phonematischen Wortflüssigkeit (PF)*. Auch der *AST* wies einen statistisch bedeutsamen Zusammenhang mit den exekutiven Funktionen der *verbalen Wortflüssigkeit (VF)* auf. Zudem zeigte sich ein statistisch hoch bedeutsamer Zusammenhang zwischen dem *AST* und dem 3. Lerndurchgang des verbalen Gedächtnisses (*DG3: Wortlistendurchgang 3*). Der *AST* korrelierte auch mit dem verbal-mnestischen Lernzuwachs (*WLG: Wortliste Gedächtnis gesamt*) und den visuokonstruktiven Fähigkeiten (*FAZ: Figuren Abzeichnen*).

9.2.1 Ergebnisse zu Fragestellung F7 und Hypothese H10

F7: Welche kognitiven Funktionen werden während der Bearbeitung des *AKGT* und *AST* beansprucht?

H10: Die Bearbeitung des *AKGT* und *AST* involviert diverse kognitive Funktionen. Es bestehen signifikante positive Korrelationen zwischen mehreren Funktionsbereichen und dem *AKGT* und *AST*.

Aus den Korrelationsanalysen geht hervor, dass die Bearbeitung des *AKGT* und des *AST* jeweils mit diversen kognitiven Funktionen zusammenhängt. Beide PVT zeigten einen statistisch hoch bedeutsamen Zusammenhang mit der kognitiven Informationsverarbeitungsgeschwindigkeit. Darüber hinaus korrelierten beide PVT mit den exekutiven Funktionsbereichen der Wortflüssigkeit. Ferner lag ein hoch signifikanter Zusammenhang zwischen dem *AKGT* und den figuralen Abrufleistungen vor, der *AST* korrelierte mit der Visuokonstruktion. Zudem lagen statistisch bedeutsame Zusammenhänge zwischen dem *AST* und der verbalen Lern- und Merkfähigkeit vor. H10 kann als bestätigt gelten.

10 Diskussion

In diesem Abschnitt werden die Ergebnisse der vorliegenden Arbeit zusammenfassend dargestellt und die Einschränkungen der Studie aufgeführt. Neben allgemeinen Kritikpunkten des methodischen Vorgehens werden die Ergebnisse und Interpretationen der einzelnen Hypothesen diskutiert. Abschließend werden Implikationen für die klinisch-praktische Diagnostik negativer Antwortverzerrungen und weitere Forschungsarbeiten abgeleitet.

10.1 Zusammenfassung der Ergebnisse

Aus den Ergebnissen der Analogstudie ging hypothesenkonform eine bessere Testperformanz der Kontrollgruppe (KG) in den PVT und neuropsychologischen Standardverfahren hervor. Dabei schätzten die Probanden der KG den Schwierigkeitsgrad diverser Aufgaben höher ein als die ES. Beide Experimentalgruppen stuften zudem den *Aggravations- und Simulationstests (AST)* einfacher ein als den *Amsterdamer Kurzzeitgedächtnistest (AKGT)* und den *Word Memory Test (WMT)*. Wider Erwarten zeigte sich jedoch kein statistisch bedeutsamer Zusammenhang zwischen der subjektiven Schwierigkeit und der Testperformanz. Anschließend wurde die experimentelle Gruppeneinteilung anhand der PVT und der neuropsychologischen Standardverfahren untersucht. Mit einem optimalen Schwellenwert des *AST* von 95.7% aus dieser Analogstudie konnten 98.5% der Probanden korrekt klassifiziert werden. Mit dem Schwellenwert des Testmanuals von 90.0% erwies sich der *AST* dagegen als weniger sensitiv im Vergleich zum *AKGT* und *WMT*.

In den verbalen Gedächtnisleistungen des *California Verbal Learning Test (CVLT)* zeigten die ES im Gegensatz zur KG deutliche Leistungseinbußen in der Lern- und Merkfähigkeit. Eine binär-logistische Regressionsanalyse mit Rückwärts- und Vorwärtsselektion und Korrektur nach Firth ergab jeweils eine Klassifikationsgenauigkeit von 98.5% mit den beiden Kennwerten *Lernsumme (LS)* und *Falsch-Positive (FP)* der Rückwärtsselektion und *Durchgang 5 (DG5)* und *Falsch-Positive (FP)* der Vorwärtsselektion. Die Regressionsanalysen mit den Untertests der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* zeigten, dass mit den *Standardabweichungen der Reaktionszeiten ohne Warnton (Ao SD)* des Untertests *Alertness* 92.4% der Probanden korrekt klassifiziert werden konnten. Auch mit den Inhibitions Kennwerten des *Go/Nogo-Paradigmas Fehlleistungen (GN F)* und *Median der Reaktionszeiten (GN Md)* sowie *Fehlleistungen (GN F)* und *Standardabweichung der Reaktionszeiten (GN SD)* lag die Klassifikationsgenauigkeit bei 92.4%. Von den Kennwerten

der *geteilten Aufmerksamkeit* eigneten sich v.a. die Parameter *auditive Auslassungen (GAa A)*, *Median der Reaktionszeiten auditiv (GAa Md)* und *Standardabweichungen der Reaktionszeiten visuell (GAv SD)* zur Gruppenvorhersage (93.9%). Insgesamt zeigte sich, dass v.a. mit einer Kombination aus *CVLT*- und *TAP*-Kennwerten die zuverlässigsten Gruppenvorhersagen erreicht werden konnten. Die beiden Prädiktorkombinationen *Standardabweichung der Reaktionszeiten im GoNogo-Paradigma (GN SD)* und *Lernsumme (LS)* sowie *Standardabweichung der Reaktionszeiten im GoNogo-Paradigma (GN SD)* und *Durchgang 5 (DG5)* ermöglichten jeweils eine 100%ig korrekte Gruppenvorhersage.

Zur Generalisierung der experimentellen Studienergebnisse wurde die Testperformanz der ES mit einer Patientenstichprobe mit verminderter Anstrengungsbereitschaft (PNA) verglichen. Während sich keine Gruppendifferenzen im *AKGT* und *WMT* abbildeten, erzielten die PNA-Patienten bessere Leistungen im *AST* als die ES ($d = -1.56$). Mit einer Rate an Richtig-Positiv-Entscheidungen von 30% (Schwellenwert 90%) bzw. 40% (Schwellenwert 95.7%) konnten die hohen Klassifikationsstatistiken der Analogstudie nicht repliziert werden. In den verbalen Gedächtnisleistungen und Aufmerksamkeitsfunktionen unterschied sich die Testperformanz lediglich in einer größeren Anzahl an *Fehlreaktionen (GN F)* der ES im *GoNogo-Paradigma*. Die höchste Rate an Richtig-Positiv-Klassifikationen der PNA von 100% konnte jeweils mit einer Kombination der Schwellenwerte der *Standardabweichung der Reaktionszeiten im GoNogo-Paradigma (GN SD)* und *Lernsumme (LS)* oder *Durchgang 5 (DG5)* des *CVLT* erzielt werden.

Im letzten Analyseschritt wurde die Testperformanz der Patienten mit negativen Antwortverzerrungen (PNA) im *AST* und *AKGT* mit einer weiteren Patientenstichprobe mit tatsächlichen kognitiven Beeinträchtigungen (PKB) verglichen. Dabei führte die mangelnde Unterscheidbarkeit zwischen tatsächlichen und simulierten Leistungseinbußen in den PVT zu einer Rate an Falsch-Positiv-Klassifikationen der PKB von bis zu 75%. Des Weiteren bestätigte sich die Annahme, dass bei der Bearbeitung der PVT diverse kognitive Funktionen beansprucht werden. Beide PVT korrelierten hoch signifikant mit der zentralnervösen Informationsverarbeitungsgeschwindigkeit. Zudem zeigten beide PVT einen statistisch bedeutsamen Zusammenhang mit dem exekutiven Funktionsbereich der Wortflüssigkeit. Der *AKGT* korrelierte mit der phonematischen Wortflüssigkeit, der *AST* mit der semantischen Wortflüssigkeit. Darüber hinaus lag ein hoch signifikanter Zusammenhang zwischen dem *AKGT* und den figuralen Abrufleistungen vor. Der *AST* korrelierte dagegen mit den

visuokonstruktiven Funktionen und der verbalen Lern- und Merkfähigkeit, im Sinne der Lernmenge insgesamt und dem Lernzuwachs.

10.2 Generelle Einschränkungen der Studie

Als generelle Einschränkung der vorliegenden Studie ist zunächst der Stichprobenumfang zu nennen. Während die experimentellen Gruppengrößen mit 33 Probanden pro Bedingung ausreichten, um die ausgeprägten Gruppendifferenzen zu erfassen, hätten größere Fallzahlen die Verläufe der ROC-Kurven spezifiziert. Dies hätte genauere Abstufungen im Kurvenverlauf und die Bestimmung präziserer Schwellenwerte ermöglicht. Darüber hinaus sind insbesondere die geringen Fallzahlen der beiden klinischen Stichproben als kritisch zu betrachten, da kleine Gruppenunterschiede auf Grund mangelnder Teststärke keine statistische Signifikanz erreichen können. Hinsichtlich der Patienten mit negativen Antwortverzerrungen (PNA) ist diesbezüglich auf die Schwierigkeiten bzw. Unmöglichkeit einer offenen Rekrutierung zu verweisen. Ein Patient mit Simulationsabsichten ist darauf bedacht, diese zu verbergen (Merten, 2014; Slick et al., 1999). Der Stichprobenumfang war somit von der Anzahl ärztlicher Konsile, den externen Gutachtenaufträgen der Arbeitsgruppe und der demonstrierten Testperformanz der Patienten in den neuropsychologischen Untersuchungen abhängig. Demzufolge konnten von 28 Datensätzen, die über einen Zeitraum von vier Jahren erhoben wurden, lediglich zehn Patienten in die Stichprobe mit negativen Antwortverzerrungen eingeschlossen werden. Eine weitere klinische Patientenstichprobe mit uneingeschränkter Anstrengungsbereitschaft hätte zudem die Berechnung vollständiger Klassifikationsstatistiken des *AST* und der ermittelten eingebetteten Parameter ermöglicht.

Auch eine größere Anzahl an Patienten mit tatsächlichen kognitiven Beeinträchtigungen (PKB) hätte die Teststärke zur Analyse der Unterschiede zwischen simulierten und tatsächlichen kognitiven Dysfunktionen verbessert. Da primär Klinikpatienten in die Studie eingeschlossen wurden, war die Fallzahl von den stationären Belegzahlen abhängig. Zudem fand zeitgleich eine weitere Studie mit Demenzpatienten statt, so dass die Mehrzahl der Patienten nicht zur Studienteilnahme beider Untersuchungen motivierbar war. Demzufolge wurden weitere Patienten im Neurozentrum Starnberg rekrutiert, wobei die Bereitschaft zur Studienteilnahme im ambulanten Rahmen deutlich geringer war als im stationären Kontext. Demnach erschien ein Stichprobenumfang von $N = 12$ aus pragmatisch-ökonomischen Gründen noch akzeptabel.

Als weiterer Kritikpunkt sind mögliche Versuchsleiter-Effekte zu nennen, da die Daten ausschließlich von der Doktorandin erhoben wurden. Demnach können Einflussfaktoren der individuellen Arbeitsweise und des Versuchsleiterverhaltens nicht ausgeschlossen werden. Durch ein standardisiertes Vorgehen und eine streng manualgetreue Anwendung der einzelnen Testverfahren wurde versucht, die Objektivität der Datenerhebung zu maximieren. Zudem ist die Besonderheit der Erhebungsbedingungen der PNA im Rahmen der Konsil- und Gutachtentätigkeit des klinischen Versorgungsauftrags der Arbeitsgruppe zu beachten. Zwar wurden die genauen Kriterien über den Einschluss der einzelnen Datensätze erst nach Abschluss aller Erhebungen festgelegt, dennoch war sich die Doktorandin über die Doppelfunktion der Erhebung im Versorgungsauftrag und eines potentiellen Studienpatienten bewusst. Die diagnostische Neutralität und Objektivität des neuropsychologischen Befunds wurde durch den fachlichen Austausch mit dem Arbeitsgruppenleiter und dessen gutachterlicher Expertise unterstützt.

Weiterhin sind potentielle Selektionseffekte in den klinischen Stichproben zu bedenken. Bei den PKB kann nicht ausgeschlossen, dass Patienten mit stärkeren kognitiven Beeinträchtigungen nicht an der Studie teilnahmen. Gründe hierfür könnten sein, dass sie bereits durch die neuropsychologische Routineuntersuchung ermüdet waren und ein weiterer Untersuchungsabschnitt zu anstrengend erschien. Folglich wären größere Leistungsdifferenzen zwischen den PKB und PNA vorstellbar. Dies hätte durch eine umfassende Dokumentation des kognitiven Status der Patienten, die eine Studienteilnahme ablehnten, kontrolliert werden können.

Ferner stellt die heterogene Zusammensetzung der PNA einen relevanten Kritikpunkt der vorliegenden Studie dar. Die PNA unterschieden sich hinsichtlich der diagnostischen Sicherheit der simulierten kognitiven Dysfunktionen sensu Slick et al. (1999). Als Einschlusskriterien galten das Vorliegen eines externalen Anreizes zur Symptomverzerrung, die abschließende neuropsychologische Beurteilung der Anstrengungsbereitschaft durch die Arbeitsgruppe und mindestens ein auffälliger Wert im *Amsterdamer Kurzzeitgedächtnistest (AKGT)* oder *Word Memory Test (WMT)*. Der *WMT* hat sich in der BV-Literatur als Goldstandard etabliert und wird bis heute in zahlreichen Untersuchungen als Außenkriterium eingesetzt (Bailey, Soble & O'Rourke, 2017; Bodenburg, 2014; Fiene et al., 2015; Meyers et al., 2014; Shura, Miskey, Rowland, Yoash-Gantz & Denning, 2016; Stevens, Bahlo, Licha, Liske & Vossler-Thies, 2016). Auch der *AKGT* gilt als gut validierter PVT. Dennoch wären weitere Indikatoren verzerrter Symptomdarstellungen, insbesondere ein *SRVT* wie

beispielsweise der *SFSS*, zur Absicherung der Diagnosestellung sicherlich von Vorteil gewesen. Da nicht von allen PNA *SFSS*-Ergebnisse vorlagen, konnte in dieser Arbeit kein kombiniertes Vorgehen aus PVT und SRVT realisiert werden. In der Patientenstichprobe erzielten 70% der PNA in beiden PVT auffällige Werte. Die Testperformanz eines Patienten lag signifikant unterhalb des Zufallsbereichs, drei weitere Patienten erzielten Ergebnisse innerhalb des Zufallsbereichs. Die Stichprobe setzte sich somit aus sicheren Simulanten ($n = 1$), wahrscheinlichen Simulanten ($n = 6$) und möglichen Simulanten ($n = 3$) zusammen. Der geringe Stichprobenumfang ermöglichte keine differenzierten Analysen der einzelnen Subgruppen. Auch in der BV-Forschung werden auf Grund der Schwierigkeiten der Zusammenstellung ausreichend großer Stichproben die Subgruppen häufig zusammengefasst (Merten et al., 2010; Whiteside et al., 2015; Wolfe et al., 2010). Bislang liegen nur vereinzelt Studien vor, die eine Differenzierung zwischen der Testperformanz von sicheren, wahrscheinlichen und möglichen Simulanten berücksichtigen. Bei Trueblood (1994) zeigte sich beispielsweise eine weitgehende Überschneidung der Testperformanz von 12 sicheren Simulanten und zehn Patienten mit fragwürdiger Anstrengungsbereitschaft. Ein alternatives Vorgehen operationalisierte beispielsweise Jones (2013) durch die Verwendung der modifizierten Slick-Kriterien nach Larrabee et al. (2007). In dieser *known-group Studie* wurden wahrscheinliche Simulanten, die anhand von auffälligen Ergebnissen in zwei gut validierten PVT diagnostiziert wurden, mit wahrscheinlichen bis sicheren Simulanten verglichen, die drei auffällige BV-Indikatoren in einer Kombination aus PVT und SRVT erzielten. Es zeigten sich keinen ausgeprägten Gruppenunterschiede.

Zudem ist die Heterogenität der Diagnosen und Medikation der beiden klinischen Stichproben PNA und PKB als weitere Einflussgrößen zu beachten. Die häufigsten Diagnosegruppen waren affektive und somatoforme Störungsbilder. Dementsprechend können störungsspezifische, genuine neuropsychologische Funktionseinschränkungen nicht ausgeschlossen werden. Depressionstypische kognitive Defizite kristallisieren sich laut einer Übersichtsarbeit von Cueni, Abbruzzese, Brühl und Herwig (2015) im episodischen Gedächtnis, in den exekutiven Funktionen (insbesondere Flexibilität, Wortflüssigkeit und Inhibitionsfähigkeit) sowie einer psychomotorischen Verlangsamung, v.a. bei älteren Patienten, heraus. In der vorliegenden Studie zeigten vier Patienten der PKB (33.3%) einen auffälligen Screeningwert in der *GDS-15*. Obwohl der moderate Zusammenhang zwischen der *GDS-15* und dem *MMSE* ($r = .552$, $p = .063$) statistisch nicht bedeutsam war, ist davon auszugehen, dass die kognitive Leistungsfähigkeit der PKB durch die depressive Symptomatik

beeinflusst wurde. Dies sollte bei der Interpretation der Beeinträchtigungen in den Exekutivfunktionen und der Informationsverarbeitungsgeschwindigkeit beachtet werden.

In der neuropsychologischen Literatur werden keine typischen kognitiven Beeinträchtigungen beschrieben, die mit somatoformen Störungsbildern assoziiert sind. Inamura et al. (2015) fanden bei Patienten ab 65 Jahren und schwer ausgeprägten somatoformen Störungsbildern Verschlechterungen des allgemeinen kognitiven Status (*MMSE*) und der attentionalen Funktionen. Inwieweit die kognitiven Dysfunktionen ein altersbezogenes oder störungsspezifisches Phänomen darstellen, wurde jedoch bislang noch nicht abschließend geklärt.

Des Weiteren ist in den beiden klinischen Stichproben der Einfluss der Medikation, insbesondere von Antidepressiva, Antipsychotika und Antidementiva, auf die kognitive Leistungsfähigkeit zu berücksichtigen. Laut Brunbauer und Laux (2013) gelten die kognitiven Nebenwirkungen von anticholinergen und sedierenden Antidepressiva in den Bereichen Psychomotorik und Gedächtnis als unumstritten. Darüber hinaus besteht nach Brunbauer und Laux (2013) Konsens darüber, dass neuere, atypische Antipsychotika kognitive Funktionen wie Exekutivfunktionen, Aufmerksamkeit und Arbeitsgedächtnis günstiger modulieren als ältere, typische Antipsychotika. Zu Antidementiva werden Acetylcholinesterasehemmer und Nootropika gezählt, die eine Progredienz leichter bis mittelschwerer Demenzen verzögern können (Brunbauer & Laux, 2013). In einer Übersichtsarbeit von Birks (2006) geht aus 13 Placebostudien und 7000 Demenzpatienten (leicht bis schwer) hervor, dass sich innerhalb von sechs Monaten das kognitive Leistungsniveau durch die Einnahme von Acetylcholinesterasehemmern verbessern ließ.

Ein weiterer Diskussionspunkt der Analogstudie stellen die Instruktionserinnerungen dar. In der BV-Literatur verweisen diverse Autoren (Bianchini et al., 2001; Goebel, 1983; Heaton et al., 1978; Rogers et al., 1993) auf eine mangelnde oder inadäquate Instruktionsumsetzung. Es ist davon auszugehen, dass sich ein klinischer Simulant während einer neuropsychologischen Untersuchung seiner Simulationsintention permanent bewusst ist. In dieser Studie wurde die Intention durch Erinnerungen an die Instruktion des Studiendesigns operationalisiert. Aus dem abschließenden Manipulationscheck ging hervor, dass die mittlere Simulationsbereitschaft der ES bei 84.82 von 100 und die Aufgabenbearbeitung entsprechend der Patientenrolle bei 86.09 von 100 lag. Während die Instruktionsumsetzung somit gewährleistet wurde, bleibt die Frage des Einflusses der Instruktionserinnerung auf die Testperformanz noch zu klären. Eine indirekte Einschätzung der Auswirkung der Instruktionserinnerung kann aus der

Testpermutation abgeleitet werden. In den Testfolgen 4 und 6 liegt eine Instruktionserinnerung zwischen Lern- und Abrufphase des *WMT*. Es wurden jedoch keine Unterschiede der Testfolge im *WMT* und im *AST* gefunden. Allerdings zeigten die ES im *AKGT* eine bessere Testperformanz in Testfolge 4 ($M = 80.40$, $SD = 5.32$) als in Testfolge 6 ($M = 58.60$, $SD = 13.37$), so dass die ES deutlich stärker simulierten, wenn der *AKGT* am Ende der Untersuchung bearbeitet wurde als am Anfang. Dies kann einen Einfluss der Instruktion am Beginn der Untersuchung, einschließlich der Fallvignette und der Ankündigung zur Instruktionserinnerungen, interpretiert werden. Inwiefern Art und Ausmaß der experimentellen Simulation auf die Instruktionserinnerungen oder die Instruktion an sich zurückzuführen sind, kann in weiterführenden Analogstudien durch einen Vergleich der Testperformanz der experimentellen Simulanten ES mit und ohne Instruktionserinnerung erfolgen.

10.3 Diskussion Hypothese H1

H1: Personen mit verminderter Anstrengungsbereitschaft erzielen in den PVT-Parametern des *AKGT*, *WMT* und *AST* signifikant schlechtere Leistungen als Personen mit maximaler Anstrengungsbereitschaft.

Hypothese H1 konnte bestätigt werden, da sich zwischen den beiden experimentellen Gruppen hochsignifikante Unterschiede abzeichneten. Die experimentellen Simulanten ES präsentierten in sämtlichen Parametern des *Amsterdamer Kurzzeitgedächtnistests (AKGT)*, *Word Memory Tests (WMT)* und *Aggravations- und Simulationstests (AST)* weit schlechtere Testleistungen als die Kontrollgruppe KG. Dabei rangierten die Effektstärken des *WMT* von $d_{WMT\ IR} = -3.71$, $d_{WMT\ DR} = -4.38$ bis $d_{WMT\ Con} = -5.08$. In der Forschungsliteratur lagen die Effektstärken in Analogstudien in einem Wertebereich von $d_{WMT\ IR} = -2.90$ bei Brockhaus und Merten (2004) bis $d_{WMT\ DR} = -6.33$ bei Merten et al. (2004). Im vorliegenden Analogdesign wurden auch im *AKGT* ausgeprägte Gruppendifferenzen ($d = -2.75$) erzielt, die ebenfalls konsistent mit sehr große Effektstärken in der BV-Literatur sind: $d = -4.61$ bei Bolan et al. (2002), $d = -3.59$ bei Blaskewitz und Merten (2006), $d = -4.11$ bei Merten und Kollegen (2004). Auch der Gruppenunterschied im *AST* der vorliegenden Arbeit ($d = -2.74$) ist mit der Effektstärke der Validierungsstudie der Testautoren ($d = -2.52$) vergleichbar.

10.4 Diskussion Hypothese H2

H2: Personen mit verminderter Anstrengungsbereitschaft erzielen in neuropsychologischen Standardverfahren signifikant schlechtere Leistungen als Personen mit maximaler Anstrengungsbereitschaft.

Da die ES erwartungsgemäß auch in sämtlichen neuropsychologischen Standardverfahren dieser Arbeit eine schlechtere Testperformanz als die KG präsentierten, wurde Hypothese H2 bestätigt. Aus der bisherigen Forschungsliteratur kristallisierten sich bislang acht Kennwerte des *California Verbal Learning Test (CVLT)* zur Erfassung negativer Antwortverzerrungen heraus: *Durchgang 5 (DG5)*, *Lernsumme (LS)*, *Verzögerte freie Wiedergabe II (VFW II)*, *Wiedergabe mit Abrufhilfe II (WAI II)*, *Trefferrate Wiedererkennen von Liste A (T)*, *Falsch Positive Wiedererkennen von Liste A (FP)*, *Diskriminationswert (d')* und *z-standardisierter Diskriminationswert (Zdw)*. Konsistent mit diesen Kennwerten, exklusive des aufwendig zu berechnenden *zDW*, wurden die Leistungsdifferenzen in den Kennwerten *DG5* ($d = -4.07$), *LS* ($d = -3.93$), d' ($d = -3.85$), *VFW I* ($d = -3.41$), *FP* ($d = -3.25$), *VFW II* ($d = -3.12$), *WA II* ($d = -2.75$) und *T* ($d = -2.34$) repliziert.

Weiter zeigten sich in den basalen und geteilten Aufmerksamkeitsleistungen sowie in der Inhibitionsfähigkeit der *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* statistisch bedeutsame Gruppenunterschiede. Die ES präsentierten in den Kennwerten dieser drei Funktionsbereiche jeweils längere Reaktionszeiten, größere Standardabweichungen der Reaktionszeiten sowie eine höhere Anfälligkeit für Reaktionsauslassungen und Fehlreaktionen. Durch die Gruppendifferenzen in den basalen Aufmerksamkeitsfunktionen (*Alertness*) wurden die Ergebnisse von Bodenburg (2014) und Fiene et al. (2015) repliziert und auf weitere Aufmerksamkeitsfunktionen generalisiert. Möglicherweise haben auch die gleiche Aufgabenpräsentation am Bildschirm und die Ergebniserfassung per Tastendruck negative Antwortverzerrungen in den geprüften Aufmerksamkeitsfunktionen begünstigt. Durch die computergestützte Erfassung entfällt die soziale Interaktion mit dem Testleiter und es ist nicht auszuschließen, dass die gefühlte Anonymität der Answerfassung zur fehlerhaften Aufgabenbearbeitung einlädt. Der Kennwert der *phasischen Alertness (Apha)* ist gesondert zu betrachten. Die ES präsentierten ausschließlich in diesem Kennwert eine bessere Testperformanz als die KG. Der Kennwert *Apha* verdeutlicht die Verkürzung der Reaktionszeit der Durchgänge mit Warnton im Vergleich zu den Durchgängen ohne Warnton. Als Erklärung kann eine stärkere Aufmerksamkeitsbindung durch den Warnton in Betracht gezogen werden und weniger ausgeprägte Antwortverzerrungen mitbedingt haben.

Des Weiteren präsentierten die ES negative Antwortverzerrungen in den Standardverfahren *Turm von London (TL-D)*, *Rey Complex Figure Test and Recognition Trial (RCFT)* und *Zahlen-Verbindungs-Test (ZVT)*. Der Einsatz dieser Verfahren in der vorliegenden Arbeit diente primär als Rahmen einer neuropsychologischen Standarduntersuchung zur Einbettung der PVT. Nach Kenntnisstand der Doktorandin liegen bislang keine empirischen Untersuchungen in der BV-Literatur zu den Verfahren *ZVT* und *TL-D* vor. Dahingegen wurde die Rey-Figur (Rey, 1941) als Diagnostikum visuokonstruktiver und figural-mnestischer Fähigkeiten in mehreren BV-Studien untersucht. Lu, Boone, Cozolino und Mitchell (2003) nehmen an, dass das Abzeichnen der Rey-Figur komplexer erscheint als dies objektiv der Fall ist und Patienten häufig unzureichende Kenntnisse über visuokonstruktive Beeinträchtigungen und den Aufbau des *RCFT* haben. Konsistent mit dem *known-group Design* von Lu et al. (2003) zeigten die ES der vorliegenden Analogstudie negative Antwortverzerrungen im Abzeichnen und im kurz verzögerten Abruf. Insgesamt verdeutlichen die Antwortverzerrungen in den additionalen Leistungstests die konsequente Umsetzung der Instruktion während der gesamten Untersuchung.

10.5 Diskussion Hypothese H3

H3: Personen mit verminderter Anstrengungsbereitschaft zeigen in subjektiv einfach erlebten PVT weniger Fehler als in schwierigeren PVT. Es besteht ein positiver Zusammenhang zwischen der subjektiven Schwierigkeit und der Anzahl an Fehlern in PVT.

Die Analyse der subjektiven Schwierigkeiten der PVT verdeutlichte, dass der *Aggravations- und Simulationstest (AST)* von beiden experimentellen Gruppen einfacher wahrgenommen wurde als der *Word Memory Test (WMT)* und der *Amsterdamer Kurzzeitgedächtnistest (AKGT)*. Allerdings konnte kein Zusammenhang zwischen der subjektiven Schwierigkeit und der Testperformanz in den PVT ermittelt werden. Die PVT-Performanz konnte ausschließlich aus der experimentellen Gruppenzugehörigkeit vorhergesagt werden. Aufgrund des Konstruktionsprinzips der verdeckten Leichtigkeit und der statistisch bedeutsamen Unterschiede im subjektiven Schwierigkeitserleben der PVT erscheint ein Einfluss der subjektiven Schwierigkeiten auf die Testperformanz sehr plausibel. Ein Kritikpunkt dieser Arbeit stellt die Erfassung der subjektiven Schwierigkeit mit einer fünfstufigen Likertskala dar, wodurch die Daten lediglich auf Ordinalskalenniveau ermittelt wurden. Es ist nicht

auszuschließen, dass mit Antworten auf Intervallskalenniveau differenziertere Schwierigkeitseinschätzungen ermittelt und mögliche Zusammenhänge zwischen Testperformanz und Testschwierigkeit abgebildet hätten werden können. In der vorliegenden Analogstudie wurden im abschließenden postexperimentellen Fragebogen mögliche Einflussvariablen wie Müdigkeit und Langeweile durch visuelle Analogskalen erfasst, während die Aufgabenschwierigkeit und die Anstrengungs- bzw. Simulationsbereitschaft anhand von Likertskalen gemessen wurde. Dieser Wechsel war in einer Verbesserung der Benutzerfreundlichkeit begründet, um durch alternierende Antwortformate die Aufmerksamkeit am Ende der Untersuchung aufrecht zu erhalten und um die inhaltlichen Unterschiede zwischen allgemeinen Einflussgrößen und der Testschwierigkeit zu verdeutlichen. Dadurch wurde jedoch die Qualität der Daten vermindert. Dies muss bei der Interpretation der Ergebnisse der subjektiven Schwierigkeiten berücksichtigt werden. Demzufolge darf auf Grund fehlender bedeutsamer Zusammenhänge nicht ausgeschlossen werden, dass die Testperformanz durch die wahrgenommene Schwierigkeit beeinflusst wird. Vielmehr kann in weiterführenden Analysen die subjektive Schwierigkeit als aufschlussreiches Kriterium zur Beurteilung neuer BV-Verfahren herangezogen werden.

10.6 Diskussion Hypothese H4

H4: Sämtliche Probanden der experimentellen Analogstudie können anhand des *AST* entsprechend ihrer Gruppenzugehörigkeit korrekt klassifiziert werden.

In der vorliegenden Arbeit wurden mit dem kritischen Schwellenwert von 90% des *Aggravations- und Simulationstests (AST)* eine Spezifität von 100% und Sensitivität von 90.91% erzielt. Somit konnte die perfekte Klassifikationsgenauigkeit von 100% der Testautoren nicht repliziert werden. Als mögliche Erklärung für die niedrigere Sensitivität kann die geringere subjektive Schwierigkeit des *AST* herangezogen werden, wodurch ein PVT an Sensitivität verliert (Guilmette et al., 1993). Durch das Heraufsetzen des Schwellenwerts auf 95.72% ließ sich die Sensitivität verbessern. Daraus resultierten für den *AST* die gleichen Klassifikationsstatistiken wie für den *Amsterdamer Kurzzeitgedächtnistest (AKGT)* und den *Word Memory Test (WMT)*.

10.7 Diskussion Hypothese H5

H5: Die experimentelle Gruppenzugehörigkeit kann zuverlässig aus den *CVLT*-Parametern vorhergesagt werden.

Die logistischen Regressionsanalysen mit den Kennwerten des *California Verbal Learning Tests (CVLT)* ergaben jeweils sehr hohe Klassifikationsstatistiken (Spezifität 97.0%, Sensitivität 100%) für die Prädiktorkombinationen *Lernsumme (LS)* und *Falsch-Positive (FP)* sowie *Durchgang 5 (DG5)* und *Falsch-Positive (FP)*. Diese Prädiktoren wurden auch in bisherigen Forschungsarbeiten ermittelt. Beispielsweise fanden Whiteside et al. (2015) zuverlässige Gruppenvorhersagen mit einer Kombination aus den Kennwerten *Falsch-Positive (FP)* und *Trefferrate (T)* und einem weiteren Aufmerksamkeitsparameter. Laut einer Untersuchung von Schutte und Kollegen (2011) ist der Kennwert *Durchgang 5 (DG5)* als vielversprechend zu betrachten. Bei Demakis (2004) wurde u.a. anderem die *Lernsumme (LS)* zur Unterscheidung zwischen optimaler und verminderter Anstrengungsbereitschaft herangezogen. Die Beurteilung der Anstrengungsbereitschaft mit dem Kennwert *Lernsumme (LS)* erscheint plausibel, da die Anzahl erinnerter Wörter über fünf Lerndurchgänge aufsummiert wird und somit auch indirekt das bewusste Zurückhalten richtig erinnerter Wörter erfasst. In einer aktuellen *known-group Studie* von Shura et al. (2016) mit Kriegsveteranen eignete sich ebenfalls der Parameter *Lernsumme (LS)* zur Vorhersage der Anstrengungsbereitschaft. Dabei fungierte der *WMT* als Außenkriterium und die Klassifikationsgenauigkeit konnte durch eine Kombination der *Lernsumme (LS)* mit dem *Rey Complex Figure Test and Recognition Trial (RCFT)* und dem *Conners' Continuous Performance Test-Second Edition* (Conners & Staff, 2004) weiter verbessert werden. Die diagnostische Güte des Kennwerts *Durchgang 5 (DG5)* lässt sich dadurch erklären, dass eine glaubhafte Präsentation kognitiver Beeinträchtigungen ein Wissen über Lernkurven voraussetzt und selbst bei kognitiven Störungen ein Lernzuwachs über mehrere Durchgänge zu beobachten ist. Es ist anzunehmen, dass diese Fachkenntnisse selten bei Probanden mit negativen Antwortverzerrungen vorliegen. Als Erklärung für die zuverlässige Gruppenvorhersage mit dem Kennwert *Falsch-Positive (FP)* kann die leichte Antwortmanipulation während der Rekognitionsprüfung in Betracht gezogen werden. Demnach braucht ein Proband mit negativen Antwortverzerrungen bei dem dichotomen Antwortformat der Rekognitionsprüfung lediglich einige Distraktoren als Wörter der ursprünglichen Wortliste zu benennen.

Insgesamt geht aus dieser Untersuchung ebenso wie aus der bisherigen BV-Literatur (Donders & Strong, 2011; Greve et al., 2009; Wolfe et al., 2010) hervor, dass sich insbesondere die Parameter der verbalen Lern- und Merkfähigkeit (*Lernsumme LS*, *Durchgang 5 DG5*) sowie der Rekognitionsfähigkeit (*Falsch-Positive FP* und *Diskriminationswert d'*) als eingebettete BV-Parameter eignen. Die Klassifikationsstatistiken der *CVLT*-Kennwerte zeichnen sich meist durch eine hohe Spezifität und eine moderate Sensitivität aus. Auch die systematische Analyse der bisherigen Untersuchungsergebnisse mit dem *Forced-Choice Wiedererkennens (FCW)* des *CVLT* zur Erfassung negativer Antwortverzerrungen von Schwartz et al. (2016) ergab bei einer Klassifikationsgenauigkeit von 72% eine hohe Spezifität von 93% und eine moderate Sensitivität von 50%. Insofern zeigten sich die Prädiktorkombinationen *Falsch-Positive (FP)* und *Lernsumme (LS)* bzw. *Durchgang 5 (DG5)* der vorliegenden Analogstudie zuverlässigere Klassifikationen (Klassifikationsgenauigkeit = 98.5%) als diverse bisherige Ergebnisse der BV-Literatur.

10.8 Diskussion Hypothese H6

H6: Die experimentelle Gruppenzugehörigkeit kann zuverlässig aus dem *TAP*-Untertest *Alertness* vorhergesagt werden. Auch die Parameter der *TAP*-Untertests der *geteilten Aufmerksamkeit* und *Inhibitionsfähigkeit* können zur Gruppenvorhersage herangezogen werden.

In der Analogstudie konnte auch mit den *TAP*-Kennwerten die Gruppenzugehörigkeit zuverlässig vorhergesagt werden. Konsistent zu den ersten BV-Studien mit Aufmerksamkeitsparametern, beispielsweise von Steck et al. (2000) sowie Willison und Tombaugh (2006), wurden gute Klassifikationsstatistiken für die *Alertness* ermittelt. Insbesondere die *Standardabweichungen der Reaktionszeiten* der *tonischen Alertness (Ao SD)* ermöglichten die Erfassung negativer Antwortverzerrungen. Dieses Ergebnis ist konsistent mit aktuellen empirischen Arbeiten (Bodenburg, 2014; Fiene et al., 2015; Stevens et al., 2016). Mit einem kritischen *T*-Wert ($Ao SD \leq 33.5$) wurde in dieser Arbeit die höchste Klassifikationsgenauigkeit von 95.5% (Sensitivität = 90.9%, Spezifität = 100%) der einzelnen *TAP*-Kennwerte ermittelt.

Zudem konnten in der vorliegenden Arbeit negative Antwortverzerrungen mit den Kennwerten der *Inhibitionsfähigkeit* erfasst werden. V.a. die Prädiktorkombinationen *Fehlreaktionen (GN F)* und *Median der Reaktionszeiten (GN Md)* sowie *Fehlreaktionen (GN*

F) und *Standardabweichung der Reaktionszeiten (GN SD)* ermöglichten jeweils zuverlässige Klassifikationen (Sensitivität = 93.9%, Spezifität = 90.9%). Dagegen schmälerte die zunehmende Aufgabenkomplexität der *geteilten Aufmerksamkeit* die diagnostische Aussagekraft der attentionalen Parameter. Das Multitasking der Dual-Task-Aufgabe der *geteilten Aufmerksamkeit* kann als steigende Aufgabenkomplexität betrachtet werden. Aus dem bisherigen Forschungsstand paralleler Informationsverarbeitung geht laut Göthe, Oberauer und Kliegl (2016) als vielfach validierte Annahme hervor, dass eine simultane Aufgabenbearbeitung in einer Verlangsamung und / oder erhöhten Fehleranfälligkeit resultiert. Diese Kosten werden mit den Annahmen des *Response-Selection-Bottleneck* (Pashler, 1984) erklärt. Das simultane Bearbeiten einfacher Wahlreaktionsaufgaben kann als dreistufige Verarbeitungskette aus Stimulusverarbeitung (sensorische Stufe), Antwortauswahl (zentrale Stufe) und Antwortausführung (motorische Stufe) beschrieben werden. Insbesondere die Antwortauswahl gilt dabei als Kapazitätsengpass (Flaschenhals). Demnach wird während des Simulierens kognitiver Dysfunktionen auf der zentralen Stufe der Antwortauswahl die Aufgabe korrekt gelöst und anschließend eine bewusste, verzerrte Antwort generiert. Die zunehmende Aufgabenkomplexität geht mit einer Abnahme der Simulationsleistung einher. Nach der Annahme des *Response-Selection-Bottleneck* wird die diagnostische Güte der eingebetteten Parameter der *TAP* primär durch die Aufgabenkomplexität determiniert und weniger durch das Antwortformat und die Einfachheit der Verzerrungsmöglichkeiten (Tastendruck, Distanz zum Testleiter).

Die zuverlässigen Gruppenvorhersagen anhand der *Alertness* und der *Inhibitionsfähigkeit* legen nahe, dass die Bearbeitung dieser *TAP*-Untertests weniger kognitive Ressourcen beansprucht als die komplexere Aufgabenstellung der *geteilten Aufmerksamkeit*. Dies spiegeln auch die subjektiven Schwierigkeiten der *TAP*-Untertests wider. Beide Gruppen schätzten die *geteilte Aufmerksamkeit* als schwierigsten Untertest und die *Alertness* als einfachsten Untertest der *TAP* ein. Insofern blieb mehr Raum für nicht-authentisches Testverhalten und Leistungsschwankungen, was sich in negativen Antwortverzerrungen zeigte. Die Ergebnisse der *TAP*-Parameter verdeutlichen, dass der Schwierigkeitsgrad der Aufgabe nicht nur bei den PVT, sondern auch bei den eingebetteten Parametern eine bedeutsame Rolle spielt. Demnach sollten die Aufgaben nicht zu viele kognitive Kapazitäten beanspruchen.

Aus den bereits erwähnten Studien von Shura et al. (2016), Schutte et al. (2011) und Whiteside et al. (2015) geht hervor, dass insbesondere die Kombination mehrerer Kennwerte

unterschiedlicher kognitiver Funktionsbereiche eine vielversprechende diagnostische Strategie darstellt. Auch in der vorliegenden Analogstudie konnten die höchsten Klassifikationsgenauigkeiten mit einer Kombination aus *CVLT*- und *TAP*-Kennwerten erzielt werden. Mit den beiden Prädiktorkombinationen *Standardabweichungen der Reaktionszeiten* im *GoNogo*-Paradigma (*GN SD*) und *Lernsumme (LS)* sowie *GN SD* und *Durchgang 5 (DG5)* lag die korrigierte Gruppenvorhersage nach Firth jeweils bei 100%. Durch die Berücksichtigung unterschiedlicher kognitiver Funktionsbereiche kann der starke Fokus auf Gedächtnisfunktionen der *PVT* erweitert werden. Dies ermöglicht eine breitere Erfassung negativer Antwortverzerrungen, sowohl hinsichtlich der kognitiven Funktionen als auch der Messzeitpunkte, wodurch die diagnostische Güte verbessert werden kann.

10.9 Diskussion Hypothese H7

H7: Es bestehen statistisch bedeutsame Unterschiede im *AST* zwischen den experimentellen Simulanten (ES) und der klinischen PNA-Stichprobe. Mit dem *AST* können weniger Richtig-Positiv-Entscheidungen der PNA getroffen werden.

Bei den Stichprobenvergleichen der experimentellen Simulanten (ES) und der Patienten mit verminderter Anstrengungsbereitschaft (PNA) sind Unterschiede in Alter, Bildungsabschluss und Psychopathologie zu berücksichtigen, insbesondere bei den nicht normierten Rohwerten der *PVT*. Die beiden Gruppen unterschieden sich jedoch ausschließlich im *AST*, nicht im *AKGT* und *WMT*. Die PNA zeigten weniger Fehler und die Verwendung des *AST* resultierte in zahlreichen Falsch-Negativ-Zuordnungen der PNA. Mit dem *AST* wurden lediglich 30% (Schwellenwert $\leq 90\%$) bzw. 40% (Schwellenwert $\leq 95.7\%$) der PNA als auffällig erkannt. Diese Ergebnisse entsprechen der Untersuchung von Merten et al. (2006), die eine unzureichende Sensitivität von 44% für den *AST* mit dem *WMT* und dem *SFSS* als Außenkriterien ermittelten. Demnach erscheint eine Generalisierung der hohen Klassifikationsstatistiken des *AST* auf den klinischen Kontext unzulässig.

Die Unterschiede zwischen den ES und PNA im *AST* verdeutlichen die Schwierigkeit, das Simulieren kognitiver Defizite experimentell zu operationalisieren. Obwohl durch das Untersuchungsdesign der vorliegenden Analogstudie das Vortäuschen kognitiver Beeinträchtigungen erfolgreich instruiert werden konnte, kann dies nicht mit einer nicht-authentischen Symptompräsentation in klinischen Untersuchungen gleichgesetzt werden. Die

tatsächlichen Anreize der Symptomverzerrungen, Krankheitsverläufe und bisherigen Untersuchungserfahrungen, Persönlichkeitsvariablen und Störungsbilder stellen eine komplexe und individuelle Konstellation an Einflussfaktoren dar, die kaum in einer experimentellen Operationalisierung zusammengefasst werden kann. Somit bleibt die Auslegung der Simulationsintention letztendlich eine individuelle Entscheidung und entzieht sich dem Einfluss des Versuchsleiters oder Kliniklers.

10.10 Diskussion Hypothese H8

H8: Es bestehen statistisch bedeutsame Unterschiede im *CVLT* und der *TAP* zwischen den experimentellen Simulanten (ES) und der klinischen PNA-Stichprobe. Mit den eingebetteten Parametern können weniger Richtig-Positiv-Entscheidungen der PNA getroffen werden.

Mit Ausnahme einer erhöhten Fehlerrate der ES im *GoNogo*-Paradigma (*GN F*) wurden zwischen den ES und PNA keine statistisch bedeutsamen Gruppenunterschiede ermittelt. Zwar präsentierten die PNA weniger ausgeprägte Leistungseinbußen in den *CVLT*-Parametern *Lernsumme (LS)*, *Durchgang 5 (DG5)* und (*kurz*) *verzögerte freie Wiedergabe I (VFW I)*, jedoch erwiesen sich diese Unterschiede nach einer Bonferroni-Holm-Korrektur nicht mehr als statistisch bedeutsam. Insgesamt deuten die Ergebnisse auf ausgeprägtere Verzerrungstendenzen bei den ES, die im Vergleich zu den PNA mehr Fehlreaktionen zeigten und weniger Wörter selbstständig reproduzierten. Wie bereits diskutiert, kann dies u.a. in den experimentellen Untersuchungsinstruktionen begründet sein. Darüber hinaus ist auch hier die geringe Fallzahl der klinischen Stichprobe zu beachten, so dass weitere Analysen mit größerer Teststärke aufschlussreich erscheinen.

Wie in der vorliegenden Arbeit wurde in der *known-group Studie* Stevens et al. (2016) neben der *TAP Alertness* auch die *Inhibitionsfähigkeit* überprüft. Patienten mit akuten Hirnläsionen ($n = 61$), Patienten in Kompensationsverfahren und Hirnläsionen ($n = 96$) sowie Patienten in Kompensationsverfahren ohne Hirnläsionen ($n = 129$) wurden in zwei Gruppen *wahrscheinliche Simulanten sensu Slick* und *keine Simulanten sensu Slick* eingeteilt. In einer schrittweisen logistischen Regressionsanalyse ermittelten die Autoren mit den Kennwerten *Standardabweichungen der Reaktionszeiten der tonischen Alertness (Ao SD)* und *Median der Reaktionszeiten der tonischen Alertness (Ao Md)* eine Klassifikationsgenauigkeit von 87%, die *Inhibitionsparameter* konnten das Vorhersagemodell jedoch nicht verbessern.

Demgegenüber erlaubten die *Standardabweichungen der Reaktionszeiten* im *GoNogo*-Paradigma (*GN SD*) in der vorliegenden Arbeit zuverlässige Richtig-Positive-Entscheidungen. Als mögliche Erklärungen können der subjektive Schwierigkeitsgrad und die angemessene Komplexität der Aufgabe in Betracht gezogen werden. Die Probanden dieser Analogstudie erlebten das *GoNogo*-Paradigma etwas anspruchsvoller als die *Alertness*, jedoch einfacher als die Aufgabenstellung der *geteilten Aufmerksamkeit*. Ferner bearbeiteten die Probanden im *GoNogo*-Paradigma nur eine Aufgabe, so dass weniger komplexe Anforderungen als in der Dualtaskaufgabe der *geteilten Aufmerksamkeit* vorlagen. Inwiefern sich die Parameter der *Inhibitionsfähigkeit* auch in einer größeren Stichprobe und einer weiteren klinischen Vergleichsgruppe zur Erfassung negativer Antwortverzerrungen eignen, erfordert weitere Untersuchungen. Wie in diversen Studien (Bodenburg, 2014; Fiene et al., 2015 & Stevens et al., 2016) eigneten sich in der vorliegenden Arbeit auch die *Alertness*-Parameter zur Erfassung negativer Antwortverzerrungen, insbesondere die *Standardabweichungen der tonischen Alertness* (*Ao SD*).

Ferner erwies sich auch in der klinischen Stichprobe der PNA die Kombination aus *CVLT*- und *TAP*-Kennwerten als diagnostisch bedeutsame Strategie zur Erfassung negativer Antwortverzerrungen. Mit den Schwellenwerten der *Standardabweichungen der Reaktionszeiten* im *GoNogo*-Paradigma (*GN SD*) und der *Lernsumme* (*LS*) oder des *Durchgang 5* (*DG5*) des *CVLT* konnten sämtliche PNA (100%) als Richtig-Positive klassifiziert werden. Auch Whiteside und Kollegen (2015) konnten in einer *known-group-Studie* und einem Vorhersagemodell mit *CVLT*-Kennwerten (*Falsch-Positive FP* und *Trefferrate T*) und einem Aufmerksamkeitsparameter des *Brief Test of Attention* (*BTA*: Schretlen et al., 1996) zufriedenstellende Klassifikationsstatistiken (Sensitivität = 54%, Spezifität = 90%) ermitteln. Allerdings ist zu bedenken, dass die Vorhersage mit den eingebetteten Parametern nicht durch eine weitere klinische Kontrollgruppe bestätigt wurde. Demzufolge kann das Risiko Falsch-Positiver-Fehlklassifikationen nicht ausgeschlossen werden und die Spezifität hat sich noch nicht im klinischen Kontext erwiesen. Vielmehr ist die Rolle der eingebetteten Parameter als ergänzendes diagnostisches Vorgehen zu PVTs zu sehen und unbedingt im Gesamtzusammenhang zu interpretieren. Jedoch können die Ergebnisse dieser Untersuchung beispielsweise als Anregung für einen PVT gesehen werden, der neben mnestischen auch attentionale Aufgaben enthält. Dabei erscheinen die Auswertung der Fehlerzahl in Gedächtnisaufgaben und auch der Standardabweichungen der

Reaktionszeiten von Aufmerksamkeitsaufgaben als sich gegenseitig ergänzende Vorgehensweisen und ermöglichen eine breitere Erfassung von Antwortverzerrungen.

10.11 Diskussion Hypothese H9 und H 10

H9: Es zeigen sich keine statistisch bedeutsamen Unterschiede zwischen den beiden Patientenstichproben im *AKGT* und *AST*. Bei den Patienten mit authentischen kognitiven Beeinträchtigungen (PKB) ist von einer erhöhten Anzahl an Falsch-Positiv-Fehlklassifikationen auszugehen.

H10: Die Bearbeitung des *AKGT* und *AST* involviert diverse kognitive Funktionen. Es bestehen signifikante positive Korrelationen zwischen mehreren Funktionsbereichen und dem *AKGT* und *AST*.

Aus den Stichprobenvergleichen zwischen den PNA und PKB gehen keine statistisch bedeutsamen Gruppenunterschiede im *AKGT* und im *AST* hervor. Neben einer generellen Verlangsamung der Informationsverarbeitungsgeschwindigkeit mit steigendem Lebensalter (Park & Minear, 2004), gehen auch demenzielle Entwicklungen mit einer Verlangsamung kognitiver Prozesse einher (Lu, Chan & Lam, 2017; Nebes & Brady, 1992; Nordlund et al., 2008). Darüber hinaus nehmen auch intraindividuelle Leistungsschwankungen mit dem Alter zu (Hultsch, Strauss, Hunter & MacDonald, 2008). In mehreren Untersuchungen wurden Standardabweichungen in die Berechnung intraindividuelle Variabilitätskoeffizienten zur Erfassung der kognitiven Entwicklung von Patienten mit beginnenden Demenzerkrankungen einbezogen (Bielak, Hultsch, Strauss, MacDonald & Hunter, 2010; Duchek et al., 2009). In der vorliegenden Studie sind die *Standardabweichungen der Reaktionszeiten (AST SD)* der PKB in den kognitiven Beeinträchtigungen begründet und übersteigen die präsentierten Reaktionszeit-Schwankungen der PNA. Mögliche Erklärungen stellen u.a. das fehlende Krankheitswissen der PNA über Art und Ausmaß der Leistungsschwankungen im Rahmen kognitiver Beeinträchtigungen dar.

Des Weiteren deuten die Ergebnisse auf eine eingeschränkte Aussagekraft des *AKGT* und des *AST* bei schweren kognitiven Beeinträchtigungen hin. Die Klassifikationsstatistiken entsprachen nicht den gängigen BV-Konventionen, da die Spezifitäten kleiner als 90% waren. Allerdings ist bei der Interpretation der Klassifikationsstatistiken die verkürzte Form des *AKGT* zu berücksichtigen. Bei drei Patienten der PKB wurde nach der Testhälfte der Schwellenwert für negative Antwortverzerrungen noch nicht erreicht, so dass die

Anstrengungsbereitschaft als unauffällig (Richtig-Negativ) betrachtet wurde. Es kann jedoch nicht ausgeschlossen werden, dass nach Beendigung der Langform der kritische Schwellenwert hätte erreicht werden können. Somit ist die Aussagekraft der Klassifikationsstatistiken des *AKGT* im Vergleich zum *AST* als eingeschränkt zu betrachten. Insgesamt ist auf Grund der hohen Anzahl an Falsch-Positiv-Klassifikationen der Einsatz von PVT bei Patienten mit schweren kognitiven Beeinträchtigungen kontraindiziert.

In dieser Studie konnte die Annahme der meisten PVT, unabhängig vom tatsächlichen kognitiven Status zu sein, nicht bestätigt werden. Vielmehr ermöglichen die Zusammenhänge der PKB zwischen den *CERADplus-NTB* Untertests und dem *AST* und dem *AKGT* Rückschlüsse auf die kognitiven Funktionsbereiche, die bei der Bearbeitung der PVT beteiligt sind. Beide PVT korrelierten hoch signifikant mit der kognitiven Informationsverarbeitungsgeschwindigkeit (*TMT A*). Auch Eberl und Wilhelm (2007) fanden einen statistisch bedeutsamen Zusammenhang zwischen dem *AST* und dem *TMT A*. Aus bisherigen Untersuchungen wie von Sanchez-Cubillo et al. (2009) geht hervor, dass die Leistung im *TMT A* insbesondere durch die visuelle Informationsverarbeitung bzw. Informationsverarbeitungsgeschwindigkeit vorhergesagt werden kann, die eine grundlegende Intelligenzfunktion darstellt. Es erscheint plausibel, dass eine schnellere Informationsverarbeitung auch eine bessere Testperformanz, einschließlich in PVT, begünstigt. Darüber hinaus bestanden signifikante Zusammenhänge zwischen den PVT und dem exekutiven Funktionsbereich der Wortflüssigkeit. Zu den Exekutivfunktionen zählen neben dem Arbeitsgedächtnis u.a. zielgerichtete Verhaltensweisen, strategische Handlungsplanung, Impulskontrolle und emotionale Selbstkontrolle sowie bewusste Aufmerksamkeitssteuerung (Müller, 2013). Bisherige Untersuchungen verdeutlichen anhand der Zusammenhänge zwischen Arbeitsgedächtnis und *AKGT* bzw. *AST* (Eberl & Wilhelm, 2007; Merten, Bossink, et al., 2007) eine Beteiligung exekutiver Funktionen bei der Bearbeitung der PVT. Eine aktuelle Studie von Kosheleva et al. (2016) mit klinisch unauffälligen Probanden ($N = 22$), die den *TOMM* bearbeiteten, verdeutlichte eine stärkere zerebrale Aktivität im superioren und medialen präfrontalen Kortex. Die exekutiven Funktionen sind im präfrontalen Kortex lokalisiert. Somit deuten auch die Ergebnisse der bildgebenden Untersuchung von Kosheleva et al. (2016) auf eine Beteiligung exekutiver Funktionen hin. In der vorliegenden Arbeit korrelierte der *AST* mit den Exekutivfunktionen der semantischen Wortflüssigkeit (*VF*), der *AKGT* mit der phonematischen Wortflüssigkeit (*PF*). Folglich konnte auch in der vorliegenden Untersuchung sowohl die Beanspruchung der

Informationsverarbeitungsgeschwindigkeit als auch exekutiver Funktionen repliziert werden. Allerdings muss bei der Interpretation der Korrelationen der moderate Zusammenhang zwischen der *GDS-15* und dem *MMSE* ($r = .552, p = .063$) berücksichtigt werden. Somit ist eine Konfundierung mit depressionstypischen kognitiven Defiziten in den exekutiven Funktionen und der psychomotorischen Verlangsamung (Cueni et al., 2015) anzunehmen. Darüber hinaus korrelierte der *AST* mit dem verbal-mnestischen Lernzuwachs (*WLG: Wortliste Gedächtnis gesamt*) und dem 3. Lerndurchgang des verbalen Gedächtnisses (*DG3: Wortlistendurchgang 3*). Der *AKGT* wies dagegen einen statistisch bedeutsamen Zusammenhang mit den figuralen Abrufleistungen (*FA: Figuren Abrufen*) auf. Diese Ergebnisse sind nicht eindeutig erklärbar. Während der *AST* neben verbalem auch bildhaft-figurales Stimulusmaterial beinhaltet, besteht der *AKGT* ausschließlich aus verbal-mnestischen Rekognitionsaufgaben. Demzufolge wäre ein Zusammenhang zwischen dem *AST* und dem figuralen Gedächtnis plausibler zu erklären als die ermittelte Korrelation zwischen dem *AKGT* und den figuralen Abrufleistungen der *CERADplus-NTB*. Auch die visuokonstruktiven Fähigkeiten (*FAZ: Figuren Abzeichnen*) korrelierten mit dem *AST*. In einer Untersuchung von Barth, Schönknecht, Pantel und Schröder (2005) zeigten Patienten mit einer leichten Demenz einzig im Untertest *Figuren Abzeichnen (FAZ)* unauffällige Ergebnisse. Als Erklärung zogen die Autoren den geringen Schwierigkeitsgrad des Untertests in Betracht, was auch hinsichtlich des Zusammenhangs mit dem *AST* erwogen werden kann. Insgesamt korrelierten die beiden *PVT* jeweils mit diversen kognitiven Funktionen. Auch in mehreren Studien mit bildgebenden Verfahren ging sowohl eine verminderte als auch unverminderte Anstrengungsbereitschaft im *WMT* und *TOMM* mit der Aktivierung komplexer neuronaler Netzwerke einher (Allen et al., 2007; Larsen et al., 2010; Spadoni et al., 2015). Zusammenfassend ist anzunehmen, dass die Bearbeitung von *PVT* eine Beteiligung unterschiedlicher kognitiver Funktionsbereiche wie exekutive Funktionen, Informationsverarbeitungsgeschwindigkeit sowie verbale und figurale Gedächtnisprozesse beansprucht. Inwieweit die Bearbeitung der *PVTs* neben der zentralnervösen Informationsverarbeitungsgeschwindigkeit auch exekutive und weitere kognitive Funktionen beansprucht werden, sollte in weiterführenden Untersuchung spezifiziert werden.

11 Implikationen für die Praxis und Ausblick

In der vorliegenden Arbeit wurden die Diskrepanzen zwischen experimentell ermittelten BV-Indikatoren und deren diagnostische Güte in klinischen Stichproben verdeutlicht. Während der *AST* im experimentellen Rahmen hohe Klassifikationsstatistiken erzielte, zeigte der *AST* eine verminderte Aussagekraft im klinischen Kontext. Der Einsatz des *AST* ist in der klinischen Praxis, auch mit einem höheren Schwellenwert von 95.7%, durch eine mangelnde Sensitivität eingeschränkt.

Darüber hinaus wurden eingebettete Parameter des *CVLT* und der *TAP* ermittelt, die zwar keinen unabhängigen PVT ersetzen, jedoch ergänzend zur Beurteilung der BV herangezogen werden können. Dabei zeigten insbesondere die Kombinationen aus *Standardabweichungen der Reaktionszeiten* im *GoNogo*-Paradigma (*GN SD*) und *Lernsumme (LS)* oder *Durchgang 5 (DG5)* des *CVLT* hohe Klassifikationsraten an Richtig-Positiv-Entscheidungen. Auf Grund der hohen Praktikabilität und Ökonomie stellen eingebettete Parameter eine weiterführende BV-Methode dar.

Zudem können aus der vorliegenden Arbeit die kognitiven Funktionsbereiche abgeleitet werden, die während Bearbeitung des *AKGT* und *AST* beansprucht werden. Das Wissen um diese Funktionsbereiche stellt eine Möglichkeit dar, das Einsatzgebiet eines PVT und dessen diagnostische Grenzen zu spezifizieren, um Fehlklassifikationen zu vermeiden. Die Auswahl und Interpretation eines PVT stellen letztendlich jedoch die diagnostischen Anforderungen eines Klinikers dar. Dabei können konkrete Einsatz- und Ausschlussempfehlungen den diagnostischen Prozess sowie die daraus resultierenden Indikationsstellungen verbessern und die praktische Tätigkeit erleichtern.

Literaturverzeichnis

- Aebi, C. (2002). *Validierung der neuropsychologischen Testbatterie CERAD-NP: eine Multi-Center Studie (Dissertation)*. Universität Basel, Philosophisch-Historische Fakultät, Basel.
- Albert, A. & Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika*, 71(1), 1-10.
- Allen, L., Conder, R., Green, P. & Cox, D. (1997). *CARB '97: Manual for the Computerized Assessment of Response Bias*. Durham, NC: CogniSyst, Inc.
- Allen, M. D., Bigler, E. D., Larsen, J., Goodrich-Hunsaker, N. J. & Hopkins, R. O. (2007). Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives. *Brain Injury*, 21(13-14), 1425-1428.
- Allgaier, A.-K., Kramer, D., Mergl, R., Fejtkova, S. & Hegerl, U. (2011). Validität der Geriatrischen Depressionsskala bei Altenheimbewohnern: Vergleich von GDS-15, GDS-8 und GDS-4. *Psychiatrische Praxis*, 38(06), 280-286.
- Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D. E. Meyer, S. Kornblum, D. E. Meyer & S. Kornblum (Hrsg.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*. (S. 183-218). Cambridge, MA, US: The MIT Press.
- Altman, D. G. (1990). *Practical statistics for medical research*. Boca Raton, FL: CRC press.
- Andreß, H.-J. (2003). *Berechnung des Konfidenzintervalls eines Anteilswerts*. Verfügbar unter: <http://eswf.uni-koeln.de/lehre/stathome/statcalc/v2202.htm>
- Armistead-Jehle, P. & Buican, B. (2013). Comparison of select Advanced Clinical Solutions embedded effort measures to the Word Memory Test in the detection of suboptimal effort. *Archives of Clinical Neuropsychology*, 28(3), 297-301.
- Audubon, J. J. & Kirwin, B. R. (1982). Defensiveness in the criminally insane. *Journal of Personality Assessment*, 46(3), 304-311.
- Bailey, K. C., Soble, J. R. & O'Rourke, J. J. F. (2017). Clinical utility of the Rey 15-Item Test, recognition trial, and error scores for detecting noncredible neuropsychological performance in a mixed clinical sample of veterans. *The Clinical Neuropsychologist*, 1-13.
- Barth, S., Schönknecht, P., Pantel, J. & Schröder, J. (2005). *Neuropsychologische Profile in der Demenzdiagnostik: Eine Untersuchung mit der CERAD-NP-Testbatterie* (Bd. 73). Basel, M. C. U. H. (2005). *CERAD-Plus test battery, revised edition* Verfügbar unter: <https://www.memoryclinic.ch/de/>
- Bashem, J. R., Rapport, L. J., Miller, J. B., Hanks, R. A., Axelrod, B. N. & Millis, S. R. (2014). Comparisons of five performance validity indices in bona fide and simulated traumatic brain injury. *The Clinical Neuropsychologist*, 28(5), 851-875.
- Bauer, L. & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: Is test security threatened? *Archives of Clinical Neuropsychology*, 21(1), 121-126.
- Bender, S. D. & Rogers, R. (2004). Detection of neurocognitive feigning: development of a multi-strategy assessment. *Archives of Clinical Neuropsychology*, 19(1), 49-60.
- Benton, A. L., Hamsher, K., Varney, N. R. & Spreen, O. (1983). *Contributions to neuropsychological assessment. A clinical manual*. New York: Oxford University Press.
- Benton, A. L. & Spreen, O. (1961). Visual memory test: The simulation of mental incompetence. *Archives of General Psychiatry*, 4(1), 79-83.

- Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology*, 12(5), 715-728.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J. & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, 27(7-8), 909-916.
- Bianchini, K. J., Curtis, K. L. & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: a dose-response relationship? *The Clinical Neuropsychologist*, 20(4), 831-847.
- Bianchini, K. J., Greve, K. W. & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *The Spine Journal*, 5(4), 404-417.
- Bianchini, K. J., Mathias, C. W. & Greve, K. W. (2001). Symptom validity testing: a critical review. *The Clinical Neuropsychologist*, 15(1), 19-45.
- Bickel, H. (2012). Das Wichtigste. Die Epidemiologie der Demenz. . Berlin: Deutsche Alzheimer Gesellschaft e.V.
- Bielak, A. A., Hultsch, D. F., Strauss, E., MacDonald, S. W. & Hunter, M. A. (2010). Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. *Neuropsychology*, 24(6), 731.
- Bigler, E. D. (2015). Neuroimaging as a biomarker in symptom validity and performance validity testing. *Brain Imaging and Behavior*, 9(3), 421-444.
- Binder, L. M. (1993). *Portland Digit Decognition Test Manual*. Portland, OR: Private Publication.
- Binder, L. M. & Rohling, M. L. (1996). Money matters: Meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry*, 153(1), 7-10.
- Birks, J. S. (2006). Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database of Systematic Reviews*(1).
- Blaskewitz, N., Gorny, I., Merten, T. & Kathmann, N. (2007). Anleitung zum nicht-authentischen Testverhalten: Coaching-Praktiken und ihre Folgen. *Praxis der Rechtspsychologie*, 17(1), 108-124.
- Blaskewitz, N. & Merten, T. (2006). Validität und Reliabilität von Beschwerdenuvalidierungstests und -indikatoren. *Zeitschrift für Neuropsychologie*, 17(1), 35-44.
- Blaskewitz, N. & Merten, T. (2007). Diagnostik der Beschwerdenuvalidität. *Fortschritte der Neurologie-Psychiatrie*, 75(03), 140-154.
- Blaskewitz, N., Merten, T. & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology*, 23(4), 379-391.
- Bodenburg, S. (2014). Die Standardabweichung der Reaktionszeiten als ein Maß für negative Antwortverzerrungen. *Zeitschrift für Neuropsychologie*, 25(2), 89-97.
- Bolan, B., Foster, J. K., Schmand, B. & Bolan, S. (2002). A comparison of three tests to detect feigned amnesia: The effects of feedback and the measurement of response latency. *Journal of Clinical and Experimental Neuropsychology*, 24(2), 154-167.
- Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach*. New York: Guilford Press.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human-und Sozialwissenschaftler. 7., vollständig überarbeitete und erweiterte Auflage*. Berlin ua: Springer.

- Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J. et al. (2014). Mechanisms of motivation–cognition interaction: Challenges and opportunities. *Cognitive, Affective & Behavioral Neuroscience*, 14(2), 443-472.
- Brickenkamp, R. (1994). *Test d2: Aufmerksamkeits-Belastungs-Test* Göttingen: Hogrefe.
- Brockhaus, R. & Merten, T. (2004). Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test. *Der Nervenarzt*, 75(9), 882-887.
- Brunnauer, A. & Laux, G. (2013). The effects of most commonly prescribed second generation antidepressants on driving ability: a systematic review. *Journal of Neural Transmission*, 120(1), 225-232.
- Brussel, J. & Hitch, K. (1943). The military malingerer. *Military Surgeon*, 93, 33-44.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H. et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, 20(4), 419-426.
- Calabrese, P. & Markowitsch, H. J. (2003). Gedächtnis und Gehirn - Neurobiologische Korrelate von Gedächtnisstörungen. *Fortschritte der Neurologie-Psychiatrie*, 71(04), 211-219.
- Cercy, S. P., Schretlen, D. J. & Brandt, J. (1997). Simulated amnesia and the pseudo-memory phenomena. In R. Rogers & R. Rogers (Hrsg.), *Clinical assessment of malingering and deception*, 2nd ed. (S. 85-107). New York: Guilford Press.
- Chafetz, M. (2011). Reducing the probability of false positives in malingering detection of Social Security disability claimants. *The Clinical Neuropsychologist*, 25(7), 1239-1252.
- Chafetz, M., Williams, M., Ben-Porath, Y., Bianchini, K., Boone, K., Kirkwood, M. et al. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, 29(6), 723-740.
- Cima, M., Merckelbach, H., Hollnack, S. & Knauer, E. (2003). Characteristics of psychiatric prison inmates who claim amnesia. *Personality and Individual Differences*, 35(2), 373-380.
- Conners, C. K. & Staff, M.-H. S. M. (2004). *Conners' Continuous Performance Test (CPT II) for Windows: Technical and software manual*. North Tonawanda, NY: Multi-Health Systems.
- Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M. & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology*, 20(2), 191-198.
- Corwin, J. & Bylsma, F. W. (1993). Psychological examination of traumatic encephalopathy. *The Clinical Neuropsychologist*, 7(1), 3-21.
- Cueni, C., Abbruzzese, E. A., Brühl, A. B. & Herwig, U. (2015). Neuropsychologische Aspekte der Depression. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*.
- Cullum, C. M., Heaton, R. K. & Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: Somatoform disorders, factitious disorders, and malingering. In H. O. Doerr, A. S. Carlin, H. O. Doerr & A. S. Carlin (Hrsg.), *Forensic neuropsychology: Legal and scientific bases*. (S. 141-171). New York: Guilford Press.
- Dandachi-FitzGerald, B., Merten, T., Ponds, R. & Niemann, H. (2015). Europäische Umfrage zum Einsatz von Beschwerdenuvalidierungstests: Ergebnisse der deutschen Teilnehmer. *Zeitschrift für Neuropsychologie*.
- Dandachi-FitzGerald, B., Ponds, R. W. & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of

- neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28(8), 771-783.
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., Peters, M. J. V. & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *The Clinical Neuropsychologist*, 25(5), 812-828.
- Dandachi-FitzGerald, B., van Twillert, B., van de Sande, P., van Os, Y. & Ponds, R. W. (2016). Poor symptom and performance validity in regularly referred Hospital outpatients: Link with standard clinical measures, and role of incentives. *Psychiatry Research*, 239, 47-53.
- Davis, J. J. & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist*, 28(2), 199-214.
- Davis, J. J. & Wall, J. R. (2014). Examining Verbal Memory on the Word Memory Test and California Verbal Learning Test. *Archives of Clinical Neuropsychology*, 29(8), 747-753.
- Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M. & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, 23(1), 133-152.
- Deister, A. (2013). Somatoforme Störungen. In H.-J. Möller, G. Laux & A. Deister (Hrsg.), *Duale Reihe Psychiatrie, Psychosomatik und Psychotherapie* (Bd. 5. Auflage). Stuttgart: Thieme.
- Delis, D., Kramer, J., Kaplan, E. & Ober, B. (1987). *California Verbal Learning Test: Adult Version*. San Antonio, TX: The Psychological Corporation.
- Delis, D., Kramer, J., Kaplan, E. & Ober, B. (2000). *The California Verbal Learning Test-Second edition*. San Antonio, TX: The Psychological Corporation.
- Demakis, G. J. (2004). Application of Clinically-Derived Malingering Cutoffs on the California Verbal Learning Test and the Wechsler Adult Intelligence Test-Revised to an Analog Malingering Study. *Applied Neuropsychology*, 11(4), 220-226.
- Denning, J. H. (2012). The efficiency and accuracy of The Test of Memory Malingering Trial 1, errors on the first 10 Items of The Test of Memory Malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology*, 27(4), 417-432.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M. & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5), 979-995.
- Dierks, C., Roßnagel, A., Hornung, G. & Jandt, S. (2009). *Rechtsgutachten zum Datenschutz in der medizinischen Forschung*. Berlin: TMF-Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
- Dilling, H. & Freyberger, H. J. (Hrsg.). (2014). *Taschenführer zur ICD-10-Klassifikation psychischer Störungen nach dem Pocket Guide von J.E. Cooper* (Bd. 7. überarbeitete Auflage entsprechend ICD-10-GM). Bern: Huber.
- Dohrenbusch, R., Henningsen, P. & Merten, T. (2011). Die Beurteilung von Aggravation und Dissimulation in der Begutachtung psychischer und psychosomatischer Störungen. *Versicherungsmedizin*, 63, 81-85.
- Donders, J. & Strong, C.-A. H. (2011). Embedded Effort Indicators on the California Verbal Learning Test-Second Edition (CVLT-II): An Attempted Cross-Validation. *The Clinical Neuropsychologist*, 25(1), 173-184.
- Donders, J. & Strong, C.-A. H. (2013). Does Green's Word Memory Test really measure memory? *Journal of Clinical and Experimental Neuropsychology*, 35(8), 827-834.
- Dressing, H., Foerster, K., Widder, B., Schneider, F. & Falkai, P. (2011). Zur Anwendung von Beschwerdvalidierungstests in der psychiatrischen Begutachtung. *Nervenarzt*, 82, 388-390.

- Dreßing, H., Widder, B. & Foerster, K. (2010). Kritische Bestandsaufnahme zum Einsatz von Beschwerdvalidierungstests in der psychiatrischen Begutachtung. *Versicherungsmedizin*, 62, 163-167.
- Duchek, J. M., Balota, D. A., Tse, C.-S., Holtzman, D. M., Fagan, A. M. & Goate, A. M. (2009). The utility of intraindividual variability in selective attention tasks as an early marker for Alzheimer's disease. *Neuropsychology*, 23(6), 746.
- Eberl, A. & Wilhelm, H. (2007). *Aggravations- und Simulationstest AST 4.0* (2. Auflage). Essen: Mnemo-Verlag.
- Ebner, G. & Kopp, H. G. (2014). Das Ganser-Syndrom – Trugbild oder Krankheit? *tellmed.ch.Psychiatrie & Neurologie*, 4.
- Engel, R. R. (Hrsg.). (2000). *MMPI-2. Minnesota Multiphasic Personality Inventory-2. Manual*. Bern: Huber.
- Engel, R. R. & Groves, J. (2012). *VEI. Verhaltens- und Erlebensinventar. Deutschsprachige Fassung des Personality Assessment Inventory (PAI) von LC Morey*. Göttingen: Hogrefe.
- Erdal, K. (2004). The effects of motivation, coaching, and knowledge of neuropsychology on the simulated malingering of head injury. *Archives of Clinical Neuropsychology*, 19(1), 73-88.
- Falkai, P. & Wittchen, H.-U. (Hrsg.). (2014). *Diagnostisches und Statistisches Manual Psychischer Störungen–DSM-5: Deutsche Ausgabe*. Göttingen: Hogrefe.
- Fastenau, P. S., Denburg, N. L. & Hufford, B. J. (1999). Adult norms for the Rey-Osterrieth Complex Figure Test and for supplemental recognition and matching trials from the Extended Complex Figure Test. *Clinical Neuropsychologist*, 13(1), 30-47.
- Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. In G. P. (Shareware) (Hrsg.), *Behavior Research Methods* (Bd. 41, S. 1149-1160): <http://www.gpower.hhu.de/>.
- Faust, D. & Ackley, M. A. (1998). Did You Think It Was Going to Be Easy? In C. R. Reynolds (Hrsg.), *Detection of Malingering during Head Injury Litigation* (S. 1-54). Boston, MA: Springer US.
- Fazio, R. L., Sanders, J. F. & Denney, R. L. (2015). Comparison of performance of the Test of Memory Malingering and Word Memory Test in a criminal forensic sample. *Archives of Clinical Neuropsychology*, 30(4), 293-301.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Thousand Oaks, CA: SAGE Publications.
- Fiene, M., Bittner, V., Fischer, J., Schwiecker, K., Heinze, H.-J. & Zaehle, T. (2015). Untersuchung der Simulationssensibilität des Alertness-Tests der Testatterie zur Aufmerksamkeitsprüfung (TAP). *Zeitschrift für Neuropsychologie*, 26(2), 73-86.
- Flaro, L., Green, P. & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury*, 21(4), 373-383.
- Folstein, M. F., Folstein, S. E., McHugh, P. R. & Fanjiang, G. (2000). *Mini-Mental State Examination: MMSE user's guide*. Odessa, FL: Psychology Assessment Resources.
- Freedman, D. & Manly, J. (2015). Use of normative data and measures of performance validity and symptom validity in assessment of cognitive function. Institute of Medicine Webpage.
- Gauggel, S. & Birkner, B. (1999). Validität und Reliabilität einer deutschen Version der Geriatrischen Depressionsskala (GDS). *Zeitschrift für Klinische Psychologie und Psychotherapie*, 28(1), 18-27.

- Gauggel, S., Schmidt, A. & Didié, M. (1994). Körperliche Beschwerden und deren Einfluss auf die Erfassung depressiver Störungen bei jüngeren und älteren Menschen. *Zeitschrift für Gerontopsychologie und -psychiatrie*, 7, 203-210.
- Gervais, R. O., Rohling, M. L., Green, P. & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology*, 19(4), 475-487.
- Ghasemi, A. & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489.
- Giebel, C., Sutcliffe, C., Verbeek, H., Zabalegui, A., Soto, M., Hallberg, I. R. et al. (2016). Depressive symptomatology and associated factors in dementia in Europe: home care versus long-term care. *International Psychogeriatrics*, 28(4), 621-630.
- Gilandas, A. J. & Touyz, S. W. (1982). The neuropsychologist as an expert witness. *Australian Psychologist*, 17(3), 255-264.
- Göbber, J., Petermann, F., Piegza, M. & Kobelt, A. (2012). Beschwerdvalidierung bei Rehabilitanden mit Migrationshintergrund in der Psychosomatik. *Die Rehabilitation*, 51(05), 356-364.
- Goebel, R. A. (1983). Detection of faking on the Halstead-Reitan neuropsychological test battery. *Journal of Clinical Psychology*, 39(5), 731-742.
- Goldhammer, F. & Hartig, J. (2012). Interpretation von Testresultaten und Testeichung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 173-201). Berlin, Heidelberg: Springer.
- Goldstein, H. (1945). A malingering key for mental tests. *Psychological Bulletin*, 42(2), 104-118.
- Gorissen, M., Sanz, J. C. & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research*, 78(2-3), 199-208.
- Gorny, I. & Merten, T. (2007). Symptom information--warning--coaching: How do they affect successful feigning in neuropsychological assessment? *Journal of Forensic Neuropsychology*, 4(4), 71-97.
- Göthe, K., Oberauer, K. & Kliegl, R. (2016). Eliminating dual-task costs by minimizing crosstalk between tasks: The role of modality and feature pairings. *Cognition*, 150, 92-108.
- Grabowsky, D. (2016). *Die schönsten Tierfabeln* (Bd.). Berlin: Bild und Heimat Verlag.
- Green, D. & Rosenfeld, B. (2011). Evaluating the gold standard: A review and meta-analysis of the Structured Interview of Reported Symptoms. *Psychological Assessment*, 23(1), 95-107.
- Green, P. (2003). Welcoming a paradigm shift in neuropsychology. *Archives of Clinical Neuropsychology*, 18(6), 625-627.
- Green, P. (2003, Revised 2005). *Manual for the Word Memory Test for Windows*. Edmonton: Green's Publishing Incorporated.
- Green, P. (2004). Testmotivation und ihre Messung : "Der Patient war anscheinend motiviert ..., doch die Testergebnisse sind ungültig". Symptomvalidierungstestung in der Neuropsychologie und der Word Memory Test. *Report Psychologie*, 29(5), 303-308.
- Green, P. (2005). *Green's Word Memory Test for Microsoft Windows: User's manual*: Green's Publications Incorporated.
- Green, P., Allen, L. & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms, US version 1.1*. Durham, NC: CogniSyst.
- Green, P. & Flaro, L. (2003). Word memory test performance in children. *Child Neuropsychology: A Journal On Normal And Abnormal Development In Childhood And Adolescence*, 9(3), 189-207.

- Green, P., Flaro, L. & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury*, 23(9), 741-750.
- Green, P., Iverson, G. L. & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, 13(10), 813-819.
- Green, P., Montijo, J. & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology*, 18(2), 86-94.
- Green, P., Rohling, M. L., Lees-Haley, P. R. & Allen, L. M., III. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, 15(12), 1045-1060.
- Greiffenstein, M. F., Baker, W. J. & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6(3), 218-224.
- Greiffenstein, M. F., Greve, K. W., Bianchini, K. J. & Baker, W. J. (2008). Test of Memory Malingering and Word Memory Test: A new comparison of failure concordance rates. *Archives of Clinical Neuropsychology*, 23(7-8), 801-807.
- Greub, B. L. & Suhr, J. A. (2006). The validity of the letter memory test as a measure of memory malingering: Robustness to coaching. *Archives of Clinical Neuropsychology*, 21(4), 249-254.
- Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A. et al. (2006). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *Neurotoxicology*, 27(6), 940-950.
- Greve, K. W., Curtis, K. L., Bianchini, K. J. & Ord, J. S. (2009). Are the Original and Second Edition of the California Verbal Learning Test Equally Accurate in Detecting Malingering? *Assessment*, 16(3), 237-248.
- Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J. & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: a comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist*, 22(5), 896-918.
- Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A. & Mattingly, M. L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology*, 22(6), 709-719.
- Gubbay, J. (o.J.). Guidelines for 1-in-5 Test. Sydney: Unpubliziertes Testmaterial.
- Guilmette, T. J., Hart, K. J. & Giuliano, A. J. (1993). Malingering detection: The use of a forced-choice method in identifying organic versus simulated memory impairment. *Clinical Neuropsychologist*, 7(1), 59-69.
- Hager, W. (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie. In G. Lüer (Hrsg.), *Allgemeine experimentelle Psychologie* (S. 43-264). Stuttgart: Fischer.
- Härting, C., Markowitsch, H., Neufeld, H., Calabrese, P., Deisinger, K. & Kessler, J. (2000). *Wechsler Gedächtnistest-Revidierte Fassung: WMS-R*. Bern: Huber.
- Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing. Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology*, 17(7), 709-714.
- Hawes, S. W. & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment*, 21(1), 112-124.
- Hawes, S. W., Boccaccini, M. T. & Murrie, D. C. (2013). Psychopathy and the combination of psychopathy and sexual deviance as predictors of sexual recidivism: Meta-analytic

- findings using the Psychopathy Checklist—Revised. *Psychological Assessment*, 25(1), 233-243.
- Heaton, R. K., Smith, H. H., Jr., Lehman, R. A. & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal Of Consulting And Clinical Psychology*, 46(5), 892-900.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J. & Millis, S. R. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23(7), 1093-1129.
- Heinze, G. & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics In Medicine*, 21(16), 2409-2419.
- Heubrock, D. & Petermann, F. (2000). *Testbatterie zur Forensischen Neuropsychologie, TBFN. Testmanual. Neuropsychologische Diagnostik bei Simulationsverdacht*. Frankfurt: Swets Test Services.
- Heubrock, D., Scholl, H. & Petermann, F. (2013). Die differentielle Validität neuropsychologischer Testverfahren zum Nachweis nicht-authentischer Störungen. *Zeitschrift für Neuropsychologie*.
- Hilsabeck, R. C., LeCompte, D. C., Marks, A. R. & Grafman, J. (2001). The Word Completion Memory Test (WCMT): A new test to detect malingered memory deficits. *Archives of Clinical Neuropsychology*, 16(7), 669-678.
- Hirsch, O. & Christiansen, H. (2015). Faking ADHD? Symptom Validity Testing and Its Relation to Self-Reported, Observer-Reported Symptoms, and Neuropsychological Measures of Attention in Adults With ADHD. *Journal Of Attention Disorders*.
- Hiscock, M. & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, 11(6), 967-974.
- Horn, W. (1969). *Prüfsystem für Schul-und Bildungsberatung: PSB; Handanweisung für die Durchführung, Auswertung und Interpretation*. Göttingen: Verlag für Psychologie.
- Hosmer, D. W. & Lemeshow, S. (2000). Interpretation of the Fitted Logistic Regression Model. In *Applied Logistic Regression* (S. 47-90). New York: John Wiley & Sons, Inc.
- Hultsch, D. F., Strauss, E., Hunter, M. A. & MacDonald, S. W. S. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik, T. A. Salthouse, F. I. M. Craik & T. A. Salthouse (Hrsg.), *The handbook of aging and cognition*, 3rd ed. (S. 491-556). New York, NY, US: Psychology Press.
- Inamura, K., Shinagawa, S., Nagata, T., Tagai, K., Nukariya, K. & Nakayama, K. (2015). Cognitive dysfunction in patients with late-life somatic symptom disorder: a comparison according to disease severity. *Psychosomatics*, 56(5), 486-494.
- Inman, T. H., Vickery, C. D., Berry, D. T. R., Lamb, D. G., Edwards, C. L. & Smith, G. T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The letter memory test. *Psychological Assessment*, 10(2), 128-139.
- Iverson, G. L., Franzen, M. D. & McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior*, 15(6), 667.
- Jager, P. (2012). Glaubhaftigkeitsbeurteilung. In H. Förstl (Hrsg.), *Theory of Mind: Neurobiologie und Psychologie sozialen Verhaltens* (S. 281-291). Berlin, Heidelberg: Springer.
- Jahn, T. & Werheid, K. (2015). *Demenzen*. Göttingen: Hogrefe.
- Jasinski, L. J., Berry, D. T. R., Shandera, A. L. & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology*, 33(3), 300-314.

- Jelicic, M., Merckelbach, H., Candel, I. & Geraerts, E. (2007). Detection of feigned cognitive dysfunction using special malingering tests: A simulation study in naïve and coached malingerers. *International Journal of Neuroscience*, 117(8), 1185-1192.
- Jones, A. (2013). Victoria Symptom Validity Test: Cutoff Scores for Psychometrically Defined Malingered Groups in a Military Sample. *The Clinical Neuropsychologist*, 27(8), 1373-1394.
- Kapur, N. (1994). The coin-in-the-hand test: a new "bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of Neurology, Neurosurgery, and Psychiatry*, 57(3), 385.
- Kemp, S., Coughlan, A. K., Rowbottom, C., Wilkinson, K., Teggart, V. & Baker, G. (2008). The base rate of effort test failure in patients with medically unexplained symptoms. *Journal of Psychosomatic Research*, 65(4), 319-325.
- Kosheleva, E., Spadoni, A. D., Strigo, I. A., Buchsbaum, M. S. & Simmons, A. N. (2016). Faking bad: The neural correlates of feigning memory impairment. *Neuropsychology*, 30(3), 377-384.
- Kummer, J. (2014). *Generator für Zufallszahlen*. Verfügbar unter: <http://rechneronline.de/zufallszahlen/>
- Larrabee, G. J. (2003). Detection of Malingering Using Atypical Performance Patterns on Standard Neuropsychological Tests. *The Clinical Neuropsychologist*, 17(3), 410-425.
- Larrabee, G. J. (2007). Identification of malingering by pattern analysis on neuropsychological tests. In G. J. Larrabee (Hrsg.), *Assessment of Malingered Neuropsychological Deficits*. (S. 80-99). New York: Oxford University Press.
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22(4), 666-679.
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18(4), 625-630.
- Larrabee, G. J. (2014). Test validity and performance validity: Considerations in providing a framework for development of an ability-focused neuropsychological test battery. *Archives of Clinical Neuropsychology*, 29(7), 695-714.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W. & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Hrsg.), *Assessment of malingered neuropsychological deficits* (S. 334-371). New York: Oxford University Press.
- Larsen, J. D., Allen, M. D., Bigler, E. D., Goodrich-Hunsaker, N. J. & Hopkins, R. O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the Word Memory Test. *Brain Injury*, 24(2), 89-99.
- Lechert, Y., Schroedter, J. & Lüttinger, P. (2006). *Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die Mikrozensus 1976–2004. ZUMA-Methodenbericht 2006/12*. Mannheim: Zentrum für Umfragen, Methoden und Analysen.
- Lees-Haley, P. R., English, L. T. & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for Personal Injury Claimants. *Psychological Reports*, 68(1), 203-210.
- Lezak, M. D. (1983). *Neuropsychological assessment*. New York: Oxford University Press.
- Lezak, M. D., Howieson, D. & Loring, D. (1995). *Neurological assessment*. New York: Oxford University.
- Lienert, G. A. (1969). *Testaufbau und Testanalyse* (3. Aufl.). Weilheim: Beltz.
- Lu, H., Chan, S. S. & Lam, L. C. (2017). 'Two-level' measurements of processing speed as cognitive markers in the differential diagnosis of DSM-5 mild neurocognitive disorders (NCD). *Scientific Reports*, 7.

- Lu, P. H., Boone, K. B., Cozolino, L. & Mitchell, C. (2003). Effectiveness of the Rey-Osterrieth Complex Figure Test and the Meyers and Meyers recognition trial in the detection of suspect effort. *The Clinical Neuropsychologist*, 17(3), 426-440.
- Mahon, J. E. (2008). Two definitions of lying. *International Journal of Applied Philosophy*, 22(2), 211-230.
- Markowitsch, H. (2002). *Dem Gedächtnis auf der Spur: Erinnern und Vergessen*. Darmstadt: Primus Verlag.
- Martin, P. K., Schroeder, R. W. & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, 29(6), 741-776.
- Martin, R. C., Franzen, M. D. & Orey, S. (1998). Magnitude of error as a strategy to detect feigned memory impairment. *Clinical Neuropsychologist*, 12(1), 84-91.
- Martins, M. & Martins, I. P. (2010). Memory malingering: Evaluating WMT criteria. *Applied Neuropsychology*, 17(3), 177-182.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939-944.
- Merten, T. (2002). Fragen der neuropsychologischen Diagnostik bei Simulationsverdacht. *Fortschritte der Neurologie - Psychiatrie*, 70(03), 126-138.
- Merten, T. (2005). Der Stellenwert der Symptomvalidierung in der neuropsychologischen Begutachtung. *Zeitschrift für Neuropsychologie*, 16(1), 29-45.
- Merten, T. (2009). Negative Antwortverzerrungen in der Begutachtung. In T. Klaus-Dieter, S. Frank, F. S. Volker GrosserA2 - Klaus-Dieter Thomann & G. Volker (Hrsg.), *Orthopädisch-unfallchirurgische Begutachtung* (S. 337-345). Munich: Urban & Fischer.
- Merten, T. (2010). Negative Antwortverzerrungen und die Untersuchung unkooperativen Verhaltens in der Begutachtung. *Der Orthopäde*, 39(3), 329-334.
- Merten, T. (2011). Beschwerdenvalidierung in der neuropsychologischen Diagnostik. *Psychologie in Österreich*, 31(2-3), 180-184.
- Merten, T. (2014). *Beschwerdenvalidierung*. Göttingen: Hogrefe.
- Merten, T., Bossink, L. & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, 29(3), 308-318.
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmand, B. A., Santamaría, P. & González-Ordi, H. (2013). Symptom validity assessment in European countries: Development and state of the art. *Clínica y Salud*, 24(3), 129-138.
- Merten, T. & Dohrenbusch, R. (2010). Testpsychologische Ansätze der Beschwerdenvalidierung. *Psychotherapeut*, 55(5), 394-400.
- Merten, T., Friedel, E., Mehren, G. & Stevens, A. (2007). Über die Validität von Persönlichkeitsprofilen in der nervenärztlichen Begutachtung. *Der Nervenarzt*, 78(5), 511-520.
- Merten, T., Friedel, E. & Stevens, A. (2006). Eingeschränkte Kooperativität in der neurologisch-psychiatrischen Begutachtung: Schätzungen zur Auftretenshäufigkeit an einer Begutachtungspopulation. *Versicherungsmedizin*, 58(1), 19-21.
- Merten, T., Green, P., Henry, M., Blaskewitz, N. & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology*, 20(6), 719-726.

- Merten, T., Henry, M. & Hilsabeck, R. (2004). Symptomvalidierungstests in der neuropsychologischen Diagnostik: eine Analogstudie. *Zeitschrift für Neuropsychologie*, 15(2), 81-90.
- Merten, T., Krahl, G., Krahl, C. & Freytag, H. (2010). Prävalenz negativer Antwortverzerrungen in der berufsgenossenschaftlichen Begutachtung. *Versicherungsmedizin*, 62(3), 126-131.
- Merten, T., Merckelbach, H., Giger, P. & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*, 9(2), 102-111.
- Merten, T., Stevens, A. & Blaskewitz, N. (2007). Beschwerdenuvalidität und Begutachtung: eine Einführung. *Praxis der Rechtspsychologie*, 17(1), 7-28.
- Meyers, J. E., Galinsky, A. M. & Volbrecht, M. (1999). Malingering and mild brain injury: how low is too low. *Applied Neuropsychology*, 6(4), 208-216.
- Meyers, J. E. & Meyers, K. R. (1995). *Rey Complex Figure Test and recognition trial professional manual*. Odessa, FL: Psychological Assessment Resources.
- Meyers, J. E., Miller, R. M., Thompson, L. M., Scalese, A. M., Allred, B. C., Rupp, Z. W. et al. (2014). Using Likelihood Ratios to Detect Invalid Performance with Performance Validity Measures. *Archives of Clinical Neuropsychology*, 29(3), 224-235.
- Meyers, J. E. & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18(3), 261-276.
- Miele, A. S., Gunner, J. H., Lynch, J. K. & McCaffrey, R. J. (2011). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology*, 27(1), 10-22.
- Miller, H. & Carlidge, N. (1974). Simulation and Malingering in Relation to Injuries of the Brain and Spinal Cord. In E. H. Feiring (Hrsg.), *Brock's Injuries of the Brain and Spinal Cord and Their Coverings* (S. 638-667). Berlin, Heidelberg: Springer.
- Miller, H. A. (2001). *Miller-Forensic Assessment of Symptoms Test. Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Millis, S. R., Putnam, S. H., Adams, K. M. & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, 7(4), 463-471.
- Mitchell, A. J., Bird, V., Rizzo, M. & Meader, N. (2010). Diagnostic validity and added value of the geriatric depression scale for depression in primary care: A meta-analysis of GDS30 and GDS15. *Journal of Affective Disorders*, 125(1), 10-17.
- Mittenberg, W., Patton, C., Canyock, E. M. & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1094-1102.
- Möller, H.-J. (2013). Organische psychische Störungen. In H.-J. Möller, G. Laux & A. Deister (Hrsg.), *Duale Reihe Psychiatrie, Psychosomatik und Psychotherapie* (Bd. 5. Auflage). Stuttgart: Thieme.
- Morel, K. R. (2012). *Manual for the Morel Emotional Numbing Test for posttraumatic stress disorder*. Las Vegas, NV: Psychometric properties.
- Morey, L. C. (1991). *Personality assessment inventory (PAI)*: Wiley Online Library.
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G. et al. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9), 1159-1165.
- Müller, S. V. (2013). *Störungen der Exekutivfunktionen*. Göttingen: Hogrefe

- Nebes, R. D. & Brady, C. B. (1992). Generalized cognitive slowing and severity of dementia in Alzheimer's disease: Implications for the interpretation of response-time data. *Journal of Clinical and Experimental Neuropsychology*, 14(2), 317-326.
- Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A. & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 Symptom Validity Scale (FBS): Verified utility in forensic practice. *The Clinical Neuropsychologist*, 24(4), 701-724.
- Niemann, H., Sturm, W., Thöne-Otto, A. & Willmes, K. (2008). *California Verbal Learning Test - deutschsprachige Adaptation*. Frankfurt/Main: Pearson Assessment & Information.
- Nies, K. J. & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology*, 9(6), 501-552.
- Nordlund, A., Rolstad, S., Klang, O. L. A., Lind, K., Pedersen, M., Blennow, K. A. J. et al. (2008). Episodic memory and speed/attention deficits are associated with Alzheimer-typical CSF abnormalities in MCI. *Journal of the International Neuropsychological Society*, 14(4), 582-590.
- Osterrieth, P. A. (1944). Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. *Archives de Psychologie*, 30, 206-356.
- Oswald, W. D. & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT)*. Göttingen: Hogrefe.
- Pankratz, L. (1979). Symptom validity testing and symptom retraining: procedures for the assessment and treatment of functional sensory deficits. *Journal of consulting and clinical psychology*, 47(2), 409.
- Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Hrsg.), *Clinical assessment of malingering and deception*. (S. 169-192). New York: Guilford Press.
- Park, D. & Minear, M. (2004). *Cognitive aging: New directions for old theories*: Oxford University Press.
- Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 358.
- Pessoa, L. & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Frontiers In Neuroscience*, 4.
- Pflüger, M., Aebi, C. & Monsch, A. U. (2003). Klinische Untersuchungsverfahren (PSYINDEXshort). *Zeitschrift für Klinische Psychologie und Psychotherapie*, 32(1), 64-67.
- Posner, M. I. & Rafal, R. D. (1987). Cognitive theories of attention and the rehabilitation of attentional deficits. In M. J. Meier, A. L. Benton & L. Diller (Hrsg.), *Neuropsychological rehabilitation* (S. 182-201). New York: Guilford Press.
- Prado-Jean, A., Couratier, P., Druet-Cabanac, M., Nubukpo, P., Bernard-Bourzeix, L., Thomas, P. et al. (2010). Specific psychological and behavioral symptoms of depression in patients with dementia. *International Journal of Geriatric Psychiatry*, 25(10), 1065-1072.
- Prigatano, G. P., Smason, I., Lamb, D. G. & Bortz, J. J. (1997). Suspected malingering and the Digit Memory Test: A replication and extension. *Archives of Clinical Neuropsychology*, 12(7), 609-619.
- Rabin, L. A., Paolillo, E. & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, 31(3), 206-230.

- Rabitsch, S. (2014). Die CERAD-Testbatterie. Ergebnisse und ihre Auswirkungen im Patienten Alltag. *psychopraxis. neuropraxis*, 17(1), 30-33.
- Rasch, B., Friese, M., Hofmann, W. J. & Naumann, E. (2010). *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Raven, J. C. (1983). *Progressive Matrices: A perceptual test of intelligence*. London: Lewis.
- Rees, L. M., Tombaugh, T. N., Gansler, D. A. & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, 10(1), 10-20.
- Reicker, L. I. (2008). The ability of reaction time tests to detect simulation: An investigation of contextual effects and criterion scores. *Archives of Clinical Neuropsychology*, 23(4), 419-431.
- Reitan, R. M. (1992). *Trail making test : manual for administration and scoring*. Tucson, Ariz: Reitan Neuropsychology Laboratory.
- Resnick, P. J. (1988). Malingering of Posttraumatic Disorders. In R. Rogers (Hrsg.), *Clinical Assessment of Malingering and Deception*. (S. 84-103). New York: Guilford Press.
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique.(Les problems.). *Archives de Psychologie*.
- Rey, A. (1958). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.
- Rey, A. (1964). The clinical examination in psychology. *Paris: Presses Universitaires de France*.
- Richman, J., Green, P., Gervais, R., Flaro, L., Merten, T., Brockhaus, R. et al. (2006). Objective Tests of Symptom Exaggeration in Independent Medical Examinations. *Journal of Occupational and Environmental Medicine*, 48(3), 303-311.
- Rienstra, A., Klein Twennaar, M. & Schmand, B. (2013). Neuropsychological Characterization of Patients with the WMT Dementia Profile. *Archives of Clinical Neuropsychology*, 28(5), 463-475.
- Rogers, R. (1990a). Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry & the Law*, 18(3), 323-333.
- Rogers, R. (1990b). Models of feigned mental illness. *Professional Psychology: Research and Practice*, 21(3), 182-188.
- Rogers, R. (1997). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Rogers, R. & Cavanaugh, J. L. (1983). 'Nothing but the truth'... A reexamination of malingering. *Journal of Psychiatry & Law*, 11(4), 443-459.
- Rogers, R., Harrell, E. H. & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, 13(3), 255-274.
- Rogers, R., Sewell, K. W., Martin, M. A. & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2), 160-177.
- Rohling, M. L. & Demakis, G. J. (2010). Bowden, Shores, & Mathias (2006): Failure to replicate or just failure to notice. Does effort still account for more variance in neuropsychological test scores than TBI severity? *The Clinical Neuropsychologist*, 24(1), 119-136.
- Rose, F. E., Hall, S. & Szalda-petree, A. D. (1995). Portland digit recognition test-computerized: Measuring response latency improves the detection of malingering. *The Clinical Neuropsychologist*, 9(2), 124-134.
- Rosen, W. G., Mohs, R. C. & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, 141(11), 1356-1364.

- Rosenhan, D. L. (1973). On being sane in insane places. *Science*, 179(4070), 250-258.
- Rudolf, M. & Müller, J. (2012). *Multivariate Verfahren: eine praxisorientierte Einführung mit Anwendungsbeispielen in SPSS*. Göttingen: Hogrefe.
- Ruffolo, L. F., Guilmette, T. J. & Willis, G. W. (2000). Comparison of time and error rates on the trail making test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. *The Clinical Neuropsychologist*, 14(2), 223-230.
- Rüsseler, J., Brett, A., Klaue, U., Sailer, M. & Münte, T. F. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurology (Online Journal)*, 94-103.
- Sachs, L. & Hedderich, J. (2009). *Angewandte Statistik : Methodensammlung mit R*. Berlin: Springer Berlin.
- Sanchez-Cubillo, I., Perianez, J. A., Adrover-Roig, D., Rodriguez-Sanchez, J. M., Rios-Lago, M., Tirapu, J. et al. (2009). Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J Int Neuropsychol Soc*, 15(3), 438-450.
- Sauer, N. & Eich, W. (2007). Somatoforme Störungen und Funktionsstörungen. *Deutsches Ärzteblatt*, 104(1-2), 45-53.
- Schagen, S., Schmand, B., Sterke, S. d. & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology*, 19(1), 43-51.
- Schmand, B., de Sterke, S. & Lindeboom, J. (1999). *Amsterdam Short-Term Memory Test: Manual*. Leiden, The Netherlands: PITS.
- Schmand, B., Lindeboom, J., Merten, T. & Millis, S. R. (2005a). *Amsterdam Short-Term Memory Test: Manual*. Leiden, The Netherlands: PITS.
- Schmand, B., Lindeboom, J., Merten, T. & Millis, S. R. (2005b). *Amsterdamer Kurzzeitgedächtnistest*. Leiden, The Netherlands: PITS.
- Schmidt, T., Lanquillon, S. & Ullmann, U. (2011). Kontroverse zu beschwerdendvalidierungsverfahren bei der begutachtung psychischer störungen. *Forensische Psychiatrie, Psychologie, Kriminologie*, 5(3), 177-183.
- Schretlen, D., Bobholz, J. H. & Brandt, J. (1996). Development and psychometric properties of the Brief Test of Attention. *Clinical Neuropsychologist*, 10(1), 80-89.
- Schroeder, R. W. & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist*, 25(3), 437-453.
- Schuhfried, G. (1991). *Wiener Reaktionsgerät. Grundprogramm*. Mödling: Schuhfried.
- Schuhfried, G. (2004). *Wiener Testsystem (WTS)*. Mödling: Schuhfried.
- Schutte, C., Millis, S., Axelrod, B. & VanDyke, S. (2011). Derivation of a composite measure of embedded symptom validity indices. *The Clinical Neuropsychologist*, 25(3), 454-462.
- Schwartz, E. S., Erdodi, L., Rodriguez, N., Ghosh, J. J., Curtain, J. R., Flashman, L. A. et al. (2016). CVLT-II Forced Choice Recognition Trial as an Embedded Validity Indicator: A Systematic Review of the Evidence. *Journal of the International Neuropsychological Society*, 22(8), 851-858.
- Schwartz, S. M., Gramling, S. E., Kerr, K. L. & Morin, C. (1998). Evaluation of intellect and deficit specific information on the ability to fake memory deficits. *International Journal of Law and Psychiatry*, 21(3), 261-272.
- Seron, X. (2014). Lying in neuropsychology. *Neurophysiologie Clinique/Clinical Neurophysiology*, 44(4), 389-403.

- Seymour, T. L. & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology*, 22(4), 475-490.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions Of The Royal Society Of London. Series B, Biological Sciences*, 298(1089), 199-209.
- Sharland, M. J. & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology*, 22(2), 213-223.
- Sheikh, J. I. & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*, 5(1-2), 165-173.
- Sherer, M., Davis, L. C., Sander, A. M., Nick, T. G., Luo, C., Pastorek, N. et al. (2015). Factors associated with word memory test performance in persons with medically documented traumatic brain injury. *The Clinical Neuropsychologist*, 29(4), 522-541.
- Sheridan, M. R. & Flowers, K. A. (2010). Reaction times and deception-the lying constant. *International Journal of Psychological Studies*, 2(2), 41.
- Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gantz, R. E. & Denning, J. H. (2016). Embedded performance validity measures with postdeployment veterans: Cross-validation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, 23(2), 94-104.
- Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, 7(2), 268-288.
- Slick, D. J., Hopp, G., Strauss, E. & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology*, 18(6), 911-922.
- Slick, D. J., Hopp, G., Strauss, E. & Thompson, G. (1997). *The Victoria Symptom Validity Test*. Odessa, FL: Psychological Assessment Resources.
- Slick, D. J., Sherman, E. M. S. & Iverson, G. L. (1999). Diagnostic Criteria for Malingered Neurocognitive Dysfunction: Proposed Standards for Clinical Practice and Research. *The Clinical Neuropsychologist*, 13(4), 545-561.
- Slick, D. J., Tan, J. E., Strauss, E. H. & Hultsch, D. F. (2004). Detecting malingering: a survey of experts' practices. *Archives of Clinical Neuropsychology*, 19(4), 465-473.
- Snowden, M. B., Atkins, D. C., Steinman, L. E., Bell, J. F., Bryant, L. L., Copeland, C. et al. (2015). Longitudinal Association of Dementia and Depression. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*, 23(9), 897-905.
- Sollman, M. J. & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology*, 26(8), 774-789.
- Spadoni, A. D., Kosheleva, E., Buchsbaum, M. S. & Simmons, A. N. (2015). Neural correlates of malingering in mild traumatic brain injury: A positron emission tomography study. *Psychiatry Research: Neuroimaging*, 233(3), 367-372.
- Spreen, O. & Benton, A. (1963). Simulation of mental deficiency on a visual memory test. *American Journal of Mental Deficiency*.
- Steck, P., Reuter, B., Meir-Korrell, S. & Schönle, P. (2000). Zur Simulierbarkeit von neuropsychologischen Defiziten bei Reaktions- und bei Intelligenztests. *Zeitschrift für Neuropsychologie*, 11(3), 128-140.

- Stern, C. (1965). Entwicklung und die Genetik von Mustern. In *Verhandlungen der Gesellschaft Deutscher Naturforscher und Ärzte: 103. Versammlung zu Weimar vom 4. bis 9. Oktober 1964* (S. 140-148). Berlin, Heidelberg: Springer.
- Stevens, A., Bahlo, S., Licha, C., Liske, B. & Vossler-Thies, E. (2016). Reaction time as an indicator of insufficient effort: Development and validation of an embedded performance validity parameter. *Psychiatry Research*, 245, 74-82.
- Stevens, A., Friedel, E., Mehren, G. & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, 157(1-3), 191-200.
- Stevens, H. (1986). Is it organic or is it functional: Is it hysteria or malingering? *Psychiatric Clinics of North America*.
- Strauss, E., Hultsch, D. F., Hunter, M., Slick, D. J., Patry, B. & Levy-Bencheton, J. (1999). Using Intraindividual Variability to Detect Malingering in Cognitive Performance. *The Clinical Neuropsychologist*, 13(4), 420-432.
- Strauss, E., Slick, D. J., Levy-Bencheton, J., Hunter, M., MacDonald, S. W. S. & Hultsch, D. F. (2002). Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology*, 17(5), 423-444.
- Sturm, W., George, S., Hildebrandt, H., Reuther, P., Schoof-Tams, K. & Wallesch, C.-W. (2009). Leitlinie Diagnostik und Therapie von Aufmerksamkeitsstörungen. *Zeitschrift für Neuropsychologie*, 20(1), 59-67.
- Suhr, J. A. & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In G. J. Larrabee (Hrsg.), *Assessment of Malingered Neuropsychological Deficits*. (S. 131-170). New York: Oxford University Press.
- Suhr, J. A. & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Hrsg.), *Assessment of Malingered Neuropsychological Deficits*. (S. 287-311). New York: Oxford University Press.
- Sweet, J. J. & Breting, L. M. G. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology*, 4(1), 6-19.
- Sweet, J. J., Wolfe, P., Sattlberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S. et al. (2000). Further Investigation of Traumatic Brain Injury Versus Insufficient Effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology*, 15(2), 105-113.
- Tan, J. E., Slick, D. J., Strauss, E. & Hultsch, D. F. (2002). How'd They Do It? Malingering Strategies on Symptom Validity Tests. *The Clinical Neuropsychologist*, 16(4), 495-505.
- Teichner, G. & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, 19(3), 455-464.
- Thalman, B., Monsch, A. U., Schneitter, M., Bernasconi, F., Aebi, C. & Camachova-Davet, Z. (2000). The CERAD neuropsychological assessment battery (CERAD-NAB) – a minimal dataset as a common tool for German-speaking Europe. *Neurobiology of Aging*, 21 (Supplement 1), 30.
- Thode, H. C. (2002). *Testing for normality*. New York: Marcel Dekker.
- Thöne-Otto, A., George, S., Hildebrandt, H., Reuther, P., Schoof-Tams, K., Sturm, W. et al. (2010). Leitlinie zur Diagnostik und Therapie von Gedächtnisstörungen. *Zeitschrift für Neuropsychologie*, 21(4), 271-281.
- Thöne-Otto, A. & Markowitsch, H. J. (2004). *Gedächtnisstörungen nach Hirnschäden*. Göttingen: Hogrefe.

- Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. North Tonawanda, NY: Multi-Health Systems
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9(3), 260-268.
- Tombaugh, T. N. & Rees, L. (2000). *Manual for the Computerized Tests of Information Processing (CTIP)*. Ottawa, ON: Carleton University.
- Tombaugh, T. N. & Rees, L. (2008). *Computerized Test of Information Processing*. Toronto: Multi-Health System Inc.
- Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. *Journal Of Clinical And Experimental Neuropsychology*, 16(4), 597-607.
- Trueblood, W. & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology*, 15(4), 578-590.
- Tucha, O. & Lange, K. W. (2004). *Turm von London - Deutsche Version*. Göttingen: Hogrefe.
- Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67.
- Urban, D. & Mayerl, J. (2011). *Regressionsanalyse: Theorie, Technik und Anwendung*. Wiesbaden: Springer.
- van Hout, M. S. E., Schmand, B., Wekking, E. M. & Deelman, B. G. (2006). Cognitive functioning in patients with suspected chronic toxic encephalopathy: Evidence for neuropsychological disturbances after controlling for insufficient effort. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(3), 296-303.
- van Hout, M. S. E., Schmand, B., Wekking, E. M., Hageman, G. & Deelman, B. G. (2003). Suboptimal performance on neuropsychological tests in patients with suspected chronic toxic encephalopathy. *Neurotoxicology*, 24(4-5), 547-551.
- Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J. & Orey, S. A. (2001). Detection of Inadequate Effort on Neuropsychological Testing: A Meta-Analytic Review of Selected Procedures. *Archives of Clinical Neuropsychology*, 16(1), 45-73.
- Walter, J., Morris, J., Swier-Vosnos, A. & Pliskin, N. (2014). Effects of severity of dementia on a symptom validity measure. *The Clinical Neuropsychologist*, 28(7), 1197-1208.
- Walters, G. D., Berry, D. T. R., Rogers, R., Payne, J. W. & Granacher, R. P. (2009). Feigned neurocognitive deficit: Taxon or dimension? *Journal of Clinical and Experimental Neuropsychology*, 31(5), 584-593.
- Walters, G. D., White, T. W. & Greene, R. L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology*, 56(1), 111-117.
- Wasyliw, O. E., Grossman, L. S., Haywood, T. W. & Cavanaugh, J. L. (1988). The detection of malingering in criminal forensic groups: MMPI validity scales. *Journal of Personality Assessment*, 52(2), 321-333.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.
- Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV)*. San Antonio: Pearson.
- Wechsler, D. (2009). *Wechsler Memory Scale – Fourth Edition (WMS-IV)*. San Antonio, TX: Pearson.

- Weinborn, M., Woods, S. P., Nulsen, C. & Leighton, A. (2012). The effects of coaching on the Verbal and Nonverbal Medical Symptom Validity Tests. *The Clinical Neuropsychologist*, 26(5), 832-849.
- Westhoff, K. & Kluck, M.-L. (2008). Grundzüge einer Theorie entscheidungsorientierten psychologisch-diagnostischen Handelns. *Psychologische Gutachten: schreiben und beurteilen*, 227-237.
- Whiteside, D. M., Gaasedelen, O. J., Hahn-Ketter, A. E., Luu, H., Miller, M. L., Persinger, V. et al. (2015). Derivation of a cross-domain embedded performance validity measure in traumatic brain injury. *The Clinical Neuropsychologist*, 29(6), 788-803.
- Widder, B., Dertwinkel, R., Egle, U. T., Foerster, K. & Schiltenswolf, M. (2007). Leitlinie für die Begutachtung von Schmerzen. *Psychotherapeut*, 52(5), 334-346.
- Widows, M. R. & Smith, G. P. (2005). SIMS: Structured Inventory of Malingered Symptomatology professional manual. . Odessa, FL: Psychological Assessment Resources.
- Willis, P. F., Farrer, T. J. & Bigler, E. D. (2011). Are effort measures sensitive to cognitive impairment? *Military medicine*, 176(12).
- Willison, J. & Tombaugh, T. N. (2006). Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neuropsychology*, 21(1), 41-52.
- Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J. & Sweet, J. J. (2010). Effort Indicators within the California Verbal Learning Test-II (CVLT-II). *The Clinical Neuropsychologist*, 24(1), 153-168.
- Wollschläger, D. (2014). Regressionsmodelle für kategoriale Daten und Zähldaten. In *Grundlagen der Datenanalyse mit R* (S. 287-323): Springer.
- Wurzer, W. (1992). *Das posttraumatische organische Psychosyndrom: psychologische Diagnostik, Begutachtung und Behandlung*: WUV, Univ.-Verlag.
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M. et al. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of psychiatric research*, 17(1), 37-49.
- Zimmermann, P. & Fimm, B. (2002). *Testbatterie zur Aufmerksamkeitsprüfung*. Herzogenrath: Psytest.
- Zimmermann, P. & Fimm, B. (2012). *Die Testbatterie zur Aufmerksamkeitsprüfung. Version 2.3*. Herzogenrath: Psytest.

Anhang

Anhang A: Tabellen

Tabelle A-1	<i>Minimale und maximale Testperformanz der CVLT-Parameter.</i>	200
Tabelle A-2	<i>Vollständige Auflistung der CVLT-Schwellenwerte.</i>	201
Tabelle A-3	<i>Minimale und maximale Testperformanz der TAP-Parameter.</i>	204
Tabelle A-4	<i>Vollständige Auflistung der TAP-Schwellenwerte.</i>	205
Tabelle A-5	<i>Klassifikationen der ES und PNA mit CVLT und TAP-Kennwerten.....</i>	210

Anhang B: Studieninformation zur Teilnahme an der Studie

Anhang C: Einverständniserklärung zur Teilnahme an der Studie

Anhang D: Zusendung der Untersuchungsergebnisse der Studie

Anhang E: Anamnesebogen

Anhang F: Instruktionen für die Kontrollgruppe

Anhang G: Instruktionen für die experimentellen Simulanten

Anhang H: Manipulationscheck

Anhang I: Postexperimenteller Fragebogen für die Kontrollgruppe

Anhang J: Postexperimenteller Fragebogen für die experimentellen Simulanten

Anhang A: Tabellen

Tabelle A-1

Minimale und maximale Testperformanz der CVLT-Parameter.

	ES (n = 33)	KG (n = 33)
	Min - Max	Min - Max
DG1	31 - 54	43 - 80
DG5	20 - 45	38 - 66
LS	20 - 45	42 - 80
VFW I	20 - 47	40 - 68
WA I	20 - 51	35 - 65
VFW II	20 - 49	36 - 64
WA II	20 - 51	32 - 62
IAW	20 - 60	29 - 62
T	20 - 57	34 - 57
FP	27 - 54	35 - 56
d'	20 - 43	34 - 61
PRW	20 - 80	35 - 73
MRW	20 - 64	39 - 71
RRW	27 - 80	29 - 57
WK	20 - 58	31 - 71
SemC	20 - 50	20 - 59

Anmerkungen. * Quasiseparation, ** Separation.

Tabelle A-2

Vollständige Auflistung der CVLT-Schwellenwerte.

DG1									
Schwellenwert	30.00	31.50	34.50	38.00	39.50	41.50	44.50	47.00	50.00
Sensitivität	.000	.061	.182	.455	.485	.667	.758	.818	.879
1 - Spezifität	.000	.000	.000	.000	.000	.000	.061	.182	.212
DG5									
Schwellenwert	53.00	55.00	57.50	59.50	62.50	68.00	72.00	76.50	81.00
Sensitivität	.939	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.424	.485	.545	.576	.697	.758	.818	.879	1.000
DG5									
Schwellenwert	19.00	23.50	28.00	30.00	32.50	36.00	38.50	41.00	41.00
Sensitivität	.000	.394	.545	.788	.879	.939	.939	.939	.939
1 - Spezifität	.000	.000	.000	.000	.000	.000	.030	.091	.091
Schwellenwert	43.50	44.50	47.00	51.00	55.50	60.00	64.00	67.00	67.00
Sensitivität	.970	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.091	.182	.182	.242	.273	.818	.970	1.000	1.000
LS									
Schwellenwert	19.00	23.50	28.00	31.50	35.00	37.50	40.00	41.50	42.50
Sensitivität	.000	.485	.727	.758	.788	.848	.909	.939	.970
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000	.030
Schwellenwert	44.00	45.50	49.00	53.00	54.50	55.50	57.00	58.50	60.00
Sensitivität	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.061	.061	.091	.182	.242	.303	.394	.455	.515
Schwellenwert	62.00	64.00	65.50	67.50	70.00	72.00	76.50	81.00	81.00
Sensitivität	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.545	.667	.697	.788	.848	.879	.939	1.000	1.000
B									
Schwellenwert	26.00	29.50	33.00	36.50	39.50	42.00	44.50	45.50	47.50
Sensitivität	.000	.091	.121	.364	.606	.636	.697	.788	.848
1 - Spezifität	.000	.000	.000	.061	.061	.061	.061	.212	.242
Schwellenwert	50.00	52.50	55.00	59.00	64.00	67.50	71.00	74.00	74.00
Sensitivität	.879	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.303	.455	.545	.667	.788	.818	.970	1.000	1.000

VFW I										
Schwellenwert	19.00	23.50	28.00	30.00	31.50	33.00	35.00	36.50	38.50	41.50
Sensitivität	.000	.273	.455	.606	.667	.758	.818	.848	.879	.970
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000	.000	.030
Schwellenwert	44.50	46.50	49.00	53.00	57.50	61.00	64.00	67.00	69.00	
Sensitivität	.970	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
1 - Spezifität	.152	.242	.273	.364	.606	.636	.879	.970	1.000	
WA I										
Schwellenwert	19.00	23.50	28.00	30.00	32.50	34.50	36.00	38.00	39.50	41.00
Sensitivität	.000	.152	.424	.515	.636	.667	.697	.727	.758	.848
1 - Spezifität	.000	.000	.000	.000	.000	.000	.061	.061	.061	.121
Schwellenwert	43.00	46.00	48.50	50.00	52.00	54.50	58.50	63.00	66.00	
Sensitivität	.909	.909	.970	.970	1.000	1.000	1.000	1.000	1.000	
1 - Spezifität	.152	.242	.242	.364	.394	.485	.545	.909	1.000	
VFW II										
Schwellenwert	19.00	23.50	28.00	30.00	33.50	37.50	39.50	40.50	43.50	
Sensitivität	.000	.303	.727	.758	.848	.848	.879	.879	.909	
1 - Spezifität	.000	.000	.000	.000	.000	.030	.061	.182	.182	
Schwellenwert	47.00	48.50	49.50	51.50	55.00	58.50	62.00	65.00		
Sensitivität	.939	.970	1.000	1.000	1.000	1.000	1.000	1.000		
1 - Spezifität	.212	.212	.242	.333	.485	.515	.879	1.000		
WA II										
Schwellenwert	19.00	23.50	28.00	30.00	31.50	33.50	37.00	39.50	42.00	45.00
Sensitivität	.000	.152	.545	.667	.727	.727	.848	.879	.879	.909
1 - Spezifität	.000	.000	.000	.000	.000	.030	.061	.061	.152	.212
Schwellenwert	46.50	47.50	49.50	51.50	53.50	55.50	58.00	61.00	63.00	
Sensitivität	.970	.970	.970	1.000	1.000	1.000	1.000	1.000	1.000	
1 - Spezifität	.212	.242	.333	.364	.485	.545	.576	.939	1.000	
IWA										
Schwellenwert	19.00	23.50	28.00	32.00	37.50	41.00	42.50			
Sensitivität	.000	.182	.242	.364	.515	.667	.697			
1 - Spezifität	.000	.000	.000	.091	.121	.182	.182			
Schwellenwert	43.50	45.00	46.50	50.50	55.00	57.00				
Sensitivität	.758	.818	.879	.909	.970	1.000				
1 - Spezifität	.182	.182	.242	.273	.848	1.000				

T									
Schwellenwert	19.00	23.50	28.00	30.00	31.50	33.00	35.50	37.50	
Sensitivität	.000	.212	.273	.424	.455	.545	.697	.758	
1 - Spezifität	.000	.000	.000	.000	.000	.000	.030	.030	
FP									
Schwellenwert	26.00	28.00	30.00	31.50	33.00	34.50			
Sensitivität	.000	.212	.424	.576	.697	.727			
1 - Spezifität	.000	.000	.000	.000	.000	.000			
Schwellenwert	36.50	40.00	42.50	48.50	55.00	57.00			
Sensitivität	.818	.848	.939	.970	1.000	1.000			
1 - Spezifität	.030	.061	.152	.182	.970	1.000			
d'									
Schwellenwert	19.00	23.50	28.00	30.00	32.50	35.50	37.50	39.00	41.00
Sensitivität	.000	.061	.152	.455	.879	.879	.909	.939	.970
1 - Spezifität	.000	.000	.000	.000	.000	.030	.061	.061	.091
Schwellenwert	42.50	44.50	46.50	48.50	50.50	54.50	59.50	62.00	
Sensitivität	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
1 - Spezifität	.121	.152	.182	.242	.273	.303	.970	1.000	

Tabelle A-3

Minimale und maximale Testperformanz der TAP-Parameter.

	ES (n = 33)	KG (n = 33)
	<i>Min - Max</i>	<i>Min - Max</i>
TAP: Alertness		
Ao M	19 - 53	29 - 74
Ao SD	19 - 60	38 - 75
Am M	19 - 51	29 - 61
Am SD	19 - 62	37 - 68
Apha	29 - 81	32 - 70
TAP: Geteilte Aufmerksamkeit		
GAa M	19 - 54	29 - 67
GAa SD	19 - 61	26 - 65
GAa A	19 - 50	31 - 50
GAv M	19 - 62	44 - 71
GAv SD	19 - 69	39 - 72
GAv A	19 - 58	34 - 60
GA F	23 - 55	34 - 56
GA A	19 - 57	33 - 61
TAP: Go/Nogo		
GN M	19 - 64	36 - 69
GN SD	19 - 54	36 - 65
GN F	23 - 54	41 - 54
GN A	19 - 50	38 - 50

Tabelle A-4

Vollständige Auflistung der TAP-Schwellenwerte.

Ao Md								
Schwellenwert	18.00	19.50	20.50	21.50	22.50	24.50	26.50	28.00
Sensitivität	.000	.455	.576	.697	.727	.758	.788	.818
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000
Schwellenwert	29.50	31.50	36.00	39.50	40.50	41.50	42.50	43.50
Sensitivität	.879	.909	.909	.909	.909	.909	.909	.909
1 - Spezifität	.030	.061	.121	.152	.242	.303	.333	.364
Schwellenwert	44.50	46.00	48.00	49.50	50.50	51.50	52.50	53.50
Sensitivität	.939	.939	.939	.939	.939	.970	.970	1.000
1 - Spezifität	.424	.455	.515	.576	.606	.606	.636	.697
Schwellenwert	55.50	58.00	59.50	62.00	64.50	67.00	71.50	75.00
Sensitivität	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.727	.818	.848	.879	.909	.939	.970	1.000
Ao SD								
Schwellenwert	18.00	19.50	20.50	22.00	24.50	26.50	28.00	
Sensitivität	.000	.606	.697	.727	.758	.818	.879	
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	
Schwellenwert	33.50	38.50	39.50	41.00	43.00	44.50	45.50	
Sensitivität	.909	.909	.909	.939	.939	.939	.939	
1 - Spezifität	.000	.030	.061	.061	.091	.121	.182	
Schwellenwert	47.00	48.50	50.50	52.50	53.50	54.50	56.00	
Sensitivität	.939	.939	.939	.939	.939	.939	.939	
1 - Spezifität	.212	.242	.273	.333	.394	.424	.455	
Schwellenwert	57.50	59.00	61.00	62.50	63.50	65.00	66.50	
Sensitivität	.970	.970	1.000	1.000	1.000	1.000	1.000	
1 - Spezifität	.485	.545	.576	.606	.727	.758	.848	
Schwellenwert	67.50	68.50	70.50	73.50	76.00			
Sensitivität	1.000	1.000	1.000	1.000	1.000			
1 - Spezifität	.879	.909	.939	.970	1.000			

Am Md								
Schwellenwert	18.00	19.50	20.50	21.50	23.00	26.00	28.50	29.50
Sensitivität	.000	.424	.485	.515	.545	.636	.667	.667
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.030
Schwellenwert	30.50	31.50	33.50	35.50	36.50	37.50	38.50	39.50
Sensitivität	.758	.788	.788	.788	.788	.848	.848	.879
1 - Spezifität	.030	.030	.091	.121	.152	.152	.182	.273
Schwellenwert	40.50	42.00	43.50	44.50	45.50	46.50	47.50	48.50
Sensitivität	.879	.879	.909	.909	.909	.939	.939	.939
1 - Spezifität	.333	.364	.455	.485	.576	.606	.636	.667
Schwellenwert	49.50	50.50	51.50	52.50	55.00	57.50	59.50	62.00
Sensitivität	.939	.970	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.697	.697	.727	.788	.848	.879	.939	1.000
Am SD								
Schwellenwert	18.00	19.50	20.50	21.50	22.50	23.50	25.50	
Sensitivität	.000	.455	.515	.636	.667	.697	.727	
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000
Schwellenwert	29.50	33.00	35.50	38.00	39.50	42.00	44.50	
Sensitivität	.758	.788	.848	.879	.879	.909	.909	
1 - Spezifität	.000	.000	.000	.030	.061	.061	.091	
Schwellenwert	45.50	46.50	47.50	49.00	50.50	51.50	52.50	
Sensitivität	.909	.909	.939	.939	.970	.970	.970	
1 - Spezifität	.152	.182	.273	.303	.303	.333	.394	
Schwellenwert	53.50	54.50	55.50	57.00	58.50	59.50	60.50	
Sensitivität	.970	.970	.970	.970	.970	.970	.970	
1 - Spezifität	.424	.485	.606	.636	.727	.758	.788	
Schwellenwert	61.50	62.50	64.00	66.50	69.00			
Sensitivität	.970	1.000	1.000	1.000	1.000			
1 - Spezifität	.818	.848	.939	.970	1.000			

Apha								
Schwellenwert	28.00	30.00	31.50	32.50	35.00	37.50	38.50	39.50
Sensitivität	.000	.030	.061	.091	.121	.152	.152	.182
1 - Spezifität	.000	.000	.000	.030	.061	.152	.182	.182
Schwellenwert	40.50	41.50	42.50	43.50	44.50	45.50	46.50	48.00
Sensitivität	.212	.212	.212	.212	.212	.212	.242	.273
1 - Spezifität	.212	.364	.424	.455	.515	.545	.697	.788
Schwellenwert	49.50	50.50	51.50	52.50	54.00	57.00	62.00	65.50
Sensitivität	.303	.333	.394	.424	.515	.545	.576	.606
1 - Spezifität	.848	.879	.909	.909	.909	.909	.909	.939
Schwellenwert	67.00	69.00	71.50	73.50	74.50	77.50	80.50	82.00
Sensitivität	.636	.697	.697	.788	.818	.848	.879	1.000
1 - Spezifität	.970	.970	1.000	1.000	1.000	1.000	1.000	1.000
GAa A								
Schwellenwert	18.00	21.00	23.50	25.00	26.50	27.50	29.00	
Sensitivität	.000	.030	.061	.121	.152	.273	.485	
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	
Schwellenwert	30.50	32.00	35.00	41.00	45.50	48.00	51.00	
Sensitivität	.576	.667	.758	.879	.909	.909	1.000	
1 - Spezifität	.000	.030	.030	.030	.242	.273	1.000	
GAa Md								
Schwellenwert	18.00	19.50	20.50	21.50	23.50	25.50	26.50	28.00
Sensitivität	.000	.333	.364	.515	.545	.636	.727	.788
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000
Schwellenwert	30.50	34.00	36.50	37.50	38.50	40.50	42.50	43.50
Sensitivität	.818	.848	.848	.879	.879	.879	.909	.909
1 - Spezifität	.030	.030	.061	.061	.091	.152	.242	.273
Schwellenwert	44.50	45.50	47.00	48.50	49.50	50.50	51.50	52.50
Sensitivität	.909	.909	.909	.909	.939	.970	.970	.970
1 - Spezifität	.303	.333	.364	.394	.424	.455	.576	.606
Schwellenwert	53.50	54.50	55.50	57.50	59.50	62.00	65.50	68.00
Sensitivität	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1 - Spezifität	.636	.697	.818	.848	.909	.939	.970	1.000

GAa SD								
Schwellenwert	18.00	20.00	21.50	22.50	23.50	24.50	25.50	27.50
Sensitivität	.000	.182	.242	.273	.424	.545	.606	.667
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.061
Schwellenwert	31.50	34.50	35.50	37.00	38.50	39.50	41.00	42.50
Sensitivität	.727	.758	.788	.818	.818	.818	.818	.818
1 - Spezifität	.061	.061	.061	.061	.091	.121	.152	.182
Schwellenwert	44.00	46.00	47.50	48.50	49.50	50.50	51.50	53.50
Sensitivität	.848	.848	.879	.879	.879	.909	.909	.909
1 - Spezifität	.273	.333	.394	.455	.515	.515	.545	.636
Schwellenwert	55.50	56.50	58.00	59.50	60.50	62.50	64.50	66.00
Sensitivität	.909	.909	.909	.939	.970	1.000	1.000	1.000
1 - Spezifität	.667	.727	.818	.879	.909	.909	.970	1.000
GAv A								
Schwellenwert	18.00	19.50	22.50	25.50	27.00	28.50	29.50	31.00
Sensitivität	.000	.061	.121	.152	.182	.242	.364	.424
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000
Schwellenwert	33.00	35.00	36.50	39.50	42.50	43.50	44.50	46.50
Sensitivität	.485	.515	.515	.545	.636	.667	.667	.697
1 - Spezifität	.000	.030	.061	.091	.121	.121	.152	.212
Schwellenwert	48.50	49.50	53.50	57.50	59.00	61.00		
Sensitivität	.758	.848	.848	.939	1.000	1.000		
1 - Spezifität	.333	.485	.576	.667	.970	1.000		

GA v Md										
Schwellenwert	18.00	19.50	20.50	21.50	23.00	25.00	27.00			
Sensitivität	.000	.212	.303	.364	.394	.424	.455			
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000			
Schwellenwert	30.00	32.50	34.50	37.00	38.50	39.50	41.00			
Sensitivität	.485	.515	.545	.576	.636	.697	.727			
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000			
Schwellenwert	43.00	44.50	45.50	46.50	47.50	49.00	51.00			
Sensitivität	.758	.758	.758	.788	.788	.818	.848			
1 - Spezifität	.000	.030	.091	.152	.182	.182	.273			
Schwellenwert	52.50	53.50	55.00	56.50	57.50	58.50	59.50			
Sensitivität	.848	.848	.909	.939	.939	.939	.939			
1 - Spezifität	.303	.364	.424	.485	.545	.606	.667			
Schwellenwert	60.50	61.50	63.00	64.50	65.50	68.50	72.00			
Sensitivität	.939	.970	1.000	1.000	1.000	1.000	1.000			
1 - Spezifität	.727	.758	.818	.879	.909	.970	1.000			
GA v SD										
Schwellenwert	18.00	19.50	21.00	23.00	24.50	27.00	29.50	30.50	31.50	33.50
Sensitivität	.000	.212	.242	.273	.333	.364	.424	.455	.485	.515
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Schwellenwert	35.50	37.00	38.50	39.50	40.50	41.50	42.50	43.50	44.50	46.50
Sensitivität	.545	.576	.606	.667	.697	.727	.758	.788	.848	.879
1 - Spezifität	.000	.000	.000	.030	.030	.030	.030	.030	.061	.091
Schwellenwert	48.50	49.50	51.00	52.50	53.50	54.50	55.50	57.50	59.50	60.50
Sensitivität	.879	.909	.909	.909	.909	.909	.939	.970	.970	.970
1 - Spezifität	.152	.152	.182	.212	.273	.303	.333	.394	.455	.515
Schwellenwert	62.00	63.50	64.50	66.00	68.00	69.50	70.50	71.50	73.00	
Sensitivität	.970	.970	.970	.970	.970	1.000	1.000	1.000	1.000	
1 - Spezifität	.545	.606	.667	.788	.848	.879	.909	.939	1.000	
GA F										
Schwellenwert	22.00	23.50	24.50	25.50	27.00	29.00	30.50	31.50	33.00	34.50
Sensitivität	.000	.030	.061	.091	.121	.152	.242	.333	.455	.455
1 - Spezifität	.000	.000	.000	.000	.000	.000	.000	.000	.000	.121
Schwellenwert	35.50	37.00	38.50	40.00	41.50	42.50	44.50	46.50	47.50	48.50
Sensitivität	.485	.545	.667	.667	.758	.758	.758	.788	.818	.848
1 - Spezifität	.152	.152	.182	.212	.273	.364	.394	.394	.545	.576

Tabelle A-5

Klassifikationen der ES und PNA mit CVLT und TAP-Kennwerten.

	PNA (n = 10)		ES (n = 33)	
	Richtig-Positiv	Falsch-Negativ	Richtig-Positiv	Falsch-Negativ
	n (%)	n (%)	n (%)	n (%)
CVLT				
DG1 ≤ 44.5	6 (60.0)	4 (40.0)	25 (75.8)	8 (24.2)
DG5 ≤ 36.0	7 (70.0)	3 (30.0)	31 (93.9)	2 (6.1)
LS ≤ 42.5	7 (70.0)	3 (30.0)	32 (97.0)	1 (3.0)
B ≤ 44.5	8 (80.0)	2 (20.0)	23 (69.7)	10 (30.3)
VFW I ≤ 41.5	6 (60.0)	4 (40.0)	32 (97.0)	1 (3.0)
WA I ≤ 43.0	8 (80.0)	2 (20.0)	30 (90.9)	3 (9.1)
VFW II ≤ 33.5	8 (80.0)	2 (20.0)	28 (84.8)	5 (15.2)
WA II ≤ 39.5	8 (80.0)	2 (20.0)	29 (87.9)	4 (12.1)
IWA ≤ 46.5	8 (80.0)	2 (20.0)	29 (87.9)	4 (12.1)
T ≤ 39.0	7 (70.0)	3 (30.0)	28 (84.8)	5 (15.2)
FP ≤ 34.5	5 (50.0)	5 (50.0)	24 (72.7)	7 (27.3)
d' ≤ 32.5	7 (70.0)	3 (30.0)	29 (87.9)	4 (12.1)
p (ES) FP, LS ≥ .405	6 (60.0)	4 (40.0)	33 (100)	0 (0)
p (ES) FP, LS ≥ .500	7 (70.0)	3 (30.0)	32 (97.0)	1 (3.0)
p (ES) FP, LS ≥ .623	7 (70.0)	3 (30.0)	32 (97.0)	1 (3.0)
p (ES) FP, DG5 ≥ .477	7 (70.0)	3 (30.0)	33 (100)	0 (0)
p (ES) FP, DG5 ≥ .500	7 (70.0)	3 (30.0)	32 (97.0)	1 (3.0)
p (ES) FP, DG5 ≥ .656	6 (60.0)	4 (40.0)	32 (97.0)	1 (3.0)
FP ≤ 34.5 und LS ≤ 42.5	7 (70.0)	3 (30.0)	33 (100)	0 (0)
FP ≤ 34.5 und DG5 ≤ 36.0	5 (50.0)	5 (50.0)	24 (72.7)	9 (27.3)
TAP-Alertness				
Ao Md ≤ 31.5	8 (80.0)	2 (20.0)	30 (90.9)	3 (9.1)
Ao SD ≤ 33.5	8 (80.0)	2 (20.0)	30 (90.9)	3 (9.1)
Am Md ≤ 31.5	8 (80.0)	2 (20.0)	26 (78.8)	7 (21.2)
Am SD ≤ 35.5	8 (80.0)	2 (20.0)	28 (84.8)	5 (15.2)
Apha ≤ 50.5	8 (80.0)	2 (20.0)	22 (66.7)	11 (33.3)
p (ES) Ao SD ≥ .720	8 (80.0)	2 (20.0)	30 (90.9)	3 (9.1)
p (ES) Ao SD ≥ .500	8 (80.0)	2 (20.0)	31 (93.9)	2 (6.1)

TAP-Geteilte Aufmerksamkeit

GAa Md \leq 34.0	6 (60.0)	4 (40.0)	28 (84.8)	5 (15.2)
GAa SD \leq 37.0	6 (60.0)	4 (40.0)	27 (81.8)	6 (18.2)
GAa A \leq 41.0	7 (70.0)	3 (30.0)	29 (87.9)	4 (12.1)
GAv Md \leq 43.0	4 (40.0)	6 (60.0)	25 (75.8)	8 (24.2)
GAv SD \leq 46.5	7 (70.0)	3 (30.0)	29 (87.9)	4 (12.1)
GAv A \leq 35.0	5 (50.0)	5 (50.0)	17 (51.5)	16 (48.5)
GA F \leq 38.5	7 (70.0)	3 (30.0)	22 (66.7)	11 (33.3)
GA A \leq 32.5	6 (60.0)	4 (40.0)	26 (78.8)	7 (21.2)
p (ES) GA: aA, aMd, vSD \geq	7 (70.0)	3 (30.0)	31 (93.9)	2 (6.1)
p (ES) GA: aA, aMd, vSD \geq	6 (60.0)	4 (40.0)	31 (93.9)	2 (6.1)
GAa A \leq 41.0, GAa Md \leq	5 (50.0)	5 (50.0)	17 (51.5)	16 (48.5)

Go/Nogo

GN Md \leq 38.5	6 (60.0)	4 (40.0)	19 (57.6)	14 (42.4)
GN SD \leq 38.5	8 (80.0)	2 (20.0)	27 (81.8)	6 (18.2)
GN F \leq 39.5	4 (40.0)	6 (60.0)	29 (87.9)	4 (12.1)
GN A \leq 38.5	6 (60.0)	4 (40.0)	25 (75.8)	8 (24.2)
p (ES) GN: F, Md \geq .500	7 (70.0)	3 (30.0)	31 (93.9)	2 (6.1)
p (ES) GN: F, Md \geq .597	6 (60.0)	4 (40.0)	31 (93.9)	2 (6.1)
p (ES) GN: F, SD \geq .500	6 (60.0)	4 (40.0)	31 (93.9)	2 (6.1)
p (ES) GN: F, SD \geq .594	6 (60.0)	4 (40.0)	30 (90.9)	3 (9.1)
GN F \leq 39.5 und GN Md \leq	7 (70.0)	3 (30.0)	32 (97.0)	1 (3.0)
GN F \leq 39.5 und GN SD \leq	8 (80.0)	2 (20.0)	31 (93.9)	2 (6.1)

CVLT und TAP

p (ES) LS, GN SD \geq .500	9 (90.0)	1 (10.0)	33 (100)	0 (0)
p (ES) LS, GN SD \geq .548	9 (90.0)	1 (10.0)	33 (100)	0 (0)
p (ES) DG5, GN SD \geq .500	9 (90.0)	1 (10.0)	33 (100)	0 (0)
p (ES) DG5, GN SD \geq .448	9 (90.0)	1 (10.0)	33 (100)	0 (0)
GN SD \leq 38.5 und LS \leq 42.5	10 (100)	0 (0)	33 (100)	0 (0)
GN SD \leq 38.5 und DG5 \leq	10 (100)	0 (0)	33 (100)	0 (0)

Anhang B: Studieninformation zur Teilnahme an der Studie



Technische Universität München

Studieninformation zur Teilnahme an der Studie
„Eine multimethodale Erfassung von Anstrengungsbereitschaft im Rahmen neuropsychologischer Untersuchungen.“
 Stephanie Stadler

Klinikum rechts der Isar
 Anstalt des öffentlichen Rechts

Klinik und Poliklinik für Psychiatrie
 und Psychotherapie
 Univ.-Prof. Dr. med. Hans Förstl
 Direktor

Klinische und Experimentelle
 Neuropsychologie
 Prof. Dr. phil. Thomas Jahn

Ismaninger Straße 22
 D-81675 München
 Durchwahl: 089 / 4140 - 4278
 Fax: 089 / 4140 - 4888
 E-Mail: th.jahn@irz.tum.de
 www.neuropsychy.med.tum.de

Sehr geehrte(r) Studienteilnehmer(in),

vielen Dank, dass Sie sich bereit erklärt haben an der Untersuchung teilzunehmen. Die Teilnahme erfordert Ihr schriftliches Einverständnis. Dafür lesen Sie bitte dieses Informationsblatt sorgfältig durch. Fragen Sie bitte nach, wenn Sie etwas nicht verstehen oder wissen möchten.

Was möchte ich untersuchen?

Gegenstand der Studie stellt die Motivation dar, mit der neuropsychologische Verfahren bearbeitet werden. Diese wird auch Anstrengungsbereitschaft bezeichnet. Im Rahmen dieser Studie möchte ich diesbezüglich unterschiedliche Verfahren testen. Dabei werden neben der Anstrengungsbereitschaft auch die Aufmerksamkeits- und Konzentrationsfähigkeit, das Gedächtnis sowie das logische Denken untersucht.

Wie läuft die Untersuchung ab?

Die Studienteilnahme erfordert eine ca. 2 -stündige Untersuchung. Diese beinhaltet einen Fragebogen mit Angaben zu Ihrer Person (Alter, Beruf, etc.) und eine abschließende kurze Befragung, wie sie die Untersuchung erlebt haben. Der Hauptteil besteht aus einer testpsychologischen Untersuchung (Aufmerksamkeit, Gedächtnis, logisches Denken). Zu Beginn jedes Untersuchungsabschnittes erkläre ich Ihnen detailliert, wie dieser zu bearbeiten.

Bei Bedarf sind selbstverständlich Pausen im Untersuchungsablauf möglich.

Was haben Sie von der Studienteilnahme?

Mit Ihrer Studienteilnahme leisten Sie einen Beitrag zur Verbesserung der klinisch- neuropsychologischen Diagnostik. Anhand der hier eingesetzten Verfahren werden neurologische und psychiatrische Patienten untersucht.

Auf Wunsch erhalten Sie einen schriftlichen Bericht über Ihre persönlichen Untersuchungsergebnisse.

Was geschieht mit meinen Daten (Anonymität)?

Die im Rahmen dieser Studie gewonnenen Informationen werden streng vertraulich behandelt. Es gilt die Schweigepflicht, d.h. Ihre persönlichen Angaben werden nicht weitergegeben. Die wissenschaftliche Auswertung und Veröffentlichung der Studienergebnisse erfolgt ausschließlich in anonymisierter Form, dass heißt ohne Nennung Ihres Namens. Ihre Testergebnisse werden mit den Testergebnissen der anderen Probanden verrechnet und ausgewertet. Demzufolge steht nicht die Leistung einer Einzelperson im Fokus des Interesses, vielmehr geht es darum, anhand zahlreicher Personen allgemeingültige Zusammenhänge zu erkennen.

Vorstand:

Univ.-Prof. Dr. Reiner Gradinger
 (Ärztlicher Direktor, Vorsitzender)



Markus Zentler
 (Kaufmännischer Direktor)

Anette Thoke-Colberg
 (Pflegedirektorin)

Univ.-Prof. Dr. Peter Henningsen
 (Dekan)

Bankverbindung:
 Bayer. Landesbank Girozentrale
 Kto-Nr. 20 272
 BLZ 700 500 00

Anhang C: Einverständniserklärung zur Teilnahme an der Studie

Probandencode: <input style="width: 100%; height: 40px;" type="text"/>	 Klinikum rechts der Isar	 Technische Universität München
Einverständniserklärung zur Teilnahme an der Studie		Klinikum rechts der Isar Anstalt des öffentlichen Rechts
„Eine multimethodale Erfassung von Anstrengungsbereitschaft im Rahmen neuropsychologischer Untersuchungen.“ Stephanie Stadler		Klinik und Poliklinik für Psychiatrie und Psychotherapie Univ.-Prof. Dr. med. Hans Förstl Direktor
		Klinische und Experimentelle Neuropsychologie
		Prof. Dr. phil. Thomas Jahn
		Ismaninger Straße 22 D-81675 München Durchwahl: 089 / 4140 - 4278 Fax: 089 / 4140 - 4888 E-Mail: th.jahn@lrz.tum.de www.neuropsychy.med.tum.de
Hiermit willige ich, Frau/Herr _____ <div style="text-align: right;">(in Druckbuchstaben)</div>		
mit meiner Unterschrift ein und erkläre mich bereit, an der o.g. Studie teilzunehmen.		
Ich habe die Studieninformation vollständig gelesen und Sinn und Zweck der Studie verstanden. Eventuell aufgetretene Fragen wurden mir verständlich und vollständig beantwortet.		
Ich bin damit einverstanden, dass die während der Untersuchung erhobenen Daten in anonymisierter Form, d.h. ohne Nennung meines Namens, wissenschaftlich ausgewertet und veröffentlicht werden dürfen. Ich weiß, dass ich diese Einwilligung jederzeit widerrufen kann, ohne dass mir daraus persönliche Nachteile entstehen.		
München, den		
Unterschrift.....		
Vorstand: Univ.-Prof. Dr. Reinher Gradinger (Ärztlicher Direktor, Vorsitzender) Markus Zentler (Kaufmännischer Direktor) Anette Thoke-Colberg (Pflegedirektorin) Univ.-Prof. Dr. Peter Henningsen (Dekan)		
Bankverbindung: Bayer. Landesbank Girozentrale Kto-Nr. 20 272 BLZ 700 500 00		

Anhang D: Zusendung der Untersuchungsergebnisse der Studie

Probandencode:



Zusendung der Untersuchungsergebnisse der Studie

„Eine multimethodale Erfassung von Anstrengungsbereitschaft im Rahmen neuropsychologischer Untersuchungen.“
Stephanie Stadler

Klinikum rechts der Isar
Anstalt des öffentlichen Rechts

Klinik und Poliklinik für Psychiatrie
und Psychotherapie
Univ.-Prof. Dr. med. Hans Förstl
Direktor

Klinische und Experimentelle
Neuropsychologie

Prof. Dr. phil. Thomas Jahn

Ismaninger Straße 22
D-81675 München
Durchwahl: 089 / 4140 - 4278
Fax: 089 / 4140 - 4888
E-Mail: th.jahn@krz.tum.de
www.neuropsychy.med.tum.de

Hiermit beantrage ich, Frau/Herr _____
(in Druckbuchstaben)

dass mir meine individuellen Testergebnisse aus der oben genannten
Untersuchung an folgende Adresse per Post oder Mail zugeschickt werden:

München, den

Unterschrift

Vorstand:
Univ.-Prof. Dr. Reinher Gradinger
(Ärztlicher Direktor, Vorsitzender)
Markus Zentler
(Kaufmännischer Direktor)
Anette Thoke-Colberg
(Pflegedirektorin)
Univ.-Prof. Dr. Peter Henningsen
(Dekan)

Bankverbindung:
Bayer. Landesbank Girozentrale
Kto-Nr. 20 272
BLZ 700 500 00

Anhang E: Anamnesebogen

Anamnesebogen		Probandencode	
„Eine multimethodale Erfassung von Anstrengungsbereitschaft im Rahmen neuropsychologischer Untersuchungen.“ Stephanie Stadler, Dipl.-Psych.			
Code			
Geburtsdatum			
Geschlecht	(0) männlich	(1) weiblich	
Muttersprache	(1) deutsch	(2) andere:	
Erlerner Beruf			
Aktuell ausgeübter Beruf			
Familienstand	(1) ledig	(5) verheiratet, aber getrennt	
	(2) feste Partnerschaft, nicht zusammenlebend	(6) verwitwet	
	(3) feste Partnerschaft, zusammenlebend	(7) geschieden	
	(4) verheiratet		
Bildungsabschluss (Casmin)	(1) Hochschulabschluss	(6) Mittlere Reife ohne berufliche Ausbildung	
	(2) Fachhochschulabschluss	(7) Hauptschulabschluss und berufliche Ausbildung	
	(3) Fachhochschulreife/Abitur und berufliche Ausbildung	(8) Hauptschulabschluss ohne berufliche Ausbildung	
	(4) Fachhochschulreife/Abitur ohne berufliche Ausbildung	(9) Kein Abschluss	
	(5) Mittlere Reife und berufliche Ausbildung		
Litten oder leiden Sie an einer oder mehreren der folgenden Erkrankungen?	Ja	Nein	
Demenz (Alzheimer)			
Amnestisches Syndrom			
Korsakoff-Syndrom			
Schädel-Hirn-Trauma			
Schlaganfall / Hirninfarkt			
Alkohol- oder Drogenmissbrauch			
Psychische Erkrankung Welche:			
Ich nehme verschreibungspflichtige Medikamente:	Ja	Nein	
(1) Grund:			
(2) Grund:			
(3) Grund:			
(4) Grund:			
Datum der Untersuchung	Beginn der Untersuchung	Ende der Untersuchung	Dauer der Untersuchung

Anhang F: Instruktionen für die Kontrollgruppe



Liebe Studienteilnehmerin, lieber Studienteilnehmer,

die folgende Untersuchung besteht aus mehreren Testverfahren.

Diese werden im klinischen Alltag sowie in der psychologisch-medizinischen Forschung eingesetzt, um subjektiv beklagte kognitive Defizite („Ich kann mich nicht mehr so gut konzentrieren.“ oder „Ich vergesse immer mehr und kann mir Vieles nicht merken.“) objektiv messbar zu machen.

Nachdem Ihnen das jeweilige Testverfahren von mir erläutert wurde, bitte ich Sie jede der einzelnen Aufgaben so gut wie möglich zu bearbeiten.

Da mit Hilfe Ihrer Testleistung ein Vergleichswert für die bestmögliche Bearbeitung ermittelt werden soll, ist es sehr wichtig für diese Studie, dass Sie versuchen, während der gesamten Untersuchung Ihre optimale Leistung zeigen.

Bei Fragen oder Unklarheiten wenden Sie sich jederzeit bitte gerne an mich!

Nochmals herzlichen Dank für Ihre Teilnahme!

Anhang G: Instruktionen für die experimentellen Simulanten

Liebe Studienteilnehmerin, lieber Studienteilnehmer,

die folgende Untersuchung besteht aus mehreren Testverfahren.

Diese werden im klinischen Alltag sowie in der psychologisch-medizinischen Forschung eingesetzt, um subjektiv beklagte kognitive Defizite („Ich kann mich nicht mehr so gut konzentrieren.“ oder „Ich vergesse immer mehr und kann mir Vieles nicht merken.“) objektiv messbar zu machen.

In dieser Studie sollen die Tests von Ihnen jedoch nicht so gut wie möglich bearbeitet werden, vielmehr sollen Sie eine verminderte Leistungsfähigkeit simulieren. Demzufolge bitte ich Sie die Aufgaben so zu bearbeiten, als hätten Sie kognitive Defizite, d.h. eine Störung des Gedächtnisses, der Konzentration und der Aufmerksamkeit sowie der Wahrnehmung und im Erkennen von Zusammenhängen. Während der Untersuchung ist Ihr Ziel, eine Leistung glaubhaft und überzeugend zu simulieren, die mit der eines Patienten mit einer kognitiven Störung vergleichbar ist. Bevor wir mit der Untersuchung beginnen, bitte ich Sie sich folgende Situation als Hilfestellung durchzulesen und sich hineinzusetzen:

Stellen Sie sich vor, Sie sitzen im Auto an der Ampel, und ein anderer Autofahrer fährt Ihnen im langsamen Tempo hinten auf. Sie wittern eine Chance und knallen selber mit dem Kopf gegen die Scheibe, damit Sie eine ordentliche Beule am Kopf bekommen. Sie taumeln aus dem Auto, fallen zu Boden, werden in die nächste neurologische Klinik gefahren und klagen in den folgenden Tagen über Kopfschmerzen und Vergesslichkeit. Die Versicherung des Unfallgegners ist bereit, eine ziemlich hohe Summe als Schmerzensgeld zu zahlen. Sie müssen beim Neuropsychologen ihre kognitiven Defizite beweisen. Dort wird Ihnen der nachfolgende Test vorgelegt.

Da mit Hilfe Ihrer Testleistung ein Vergleichswert für Simulationsergebnisse ermittelt werden soll, ist es von großer Bedeutung für diese Studie, dass Sie versuchen während der gesamten Untersuchung glaubhaft eine schlechtere Leistung zu zeigen, als Sie es eigentlich könnten.

Bei Fragen oder Unklarheiten wenden Sie sich jederzeit bitte gerne an mich!

Nochmals herzlichen Dank für Ihre Teilnahme!

Anhang H: Manipulationscheck

Manipulationscheck	Probandencode
„Eine multimethodale Erfassung von Anstrengungsbereitschaft im Rahmen neuropsychologischer Untersuchungen.“ Stephanie Stadler, Dipl.-Psych.	
Können Sie kurz mit eigenen Worten beschreiben, was Sie während der Untersuchung tun sollen?	
ES	
während Untersuchung kognitive Defizite simulieren	
Patientenrolle	
Glaubhaftigkeit	

Anhang I: Postexperimenteller Fragebogen für die Kontrollgruppe

Postexperimenteller Fragebogen		Probandencode
KG (Testfolge 1)		
<small>Stephanie Stadler, Dipl.-Psych.</small>		
<p>Im Folgenden finden Sie einige Fragen über die soeben absolvierte Untersuchung und ihr Befinden während dieser Untersuchung. Bitte beantworten Sie die Formulierungen wahrheitsgemäß.</p> <p>Bitte geben Sie an, inwieweit eine Formulierung auf Sie zutrifft. Hierfür kreuzen Sie bitte auf dem Zahlenstrahl den entsprechenden Wert von Null (0= minimal, im Sinne von gar nicht) bis Hundert (100= maximal, im Sinne von dem stärksten Ausprägungsgrad) an.</p>		
Während der Untersuchung...	0	100
... war ich müde.	_____	
... war ich gelangweilt.	_____	
... habe ich alle Aufgaben motiviert bearbeitet.	_____	
... hatte ich Interesse an den Tests.	_____	
... stand ich der Untersuchung misstrauisch gegenüber.	_____	
... habe ich versucht die Aufgaben so gut wie möglich zu bearbeiten.	_____	

1

Anhang J: Postexperimenteller Fragebogen für die experimentellen Simulanten

Postexperimenteller Fragebogen		Probandencode
ExpSim (Testfolge 1)		
<small>Stephanie Stadler, Dipl.-Psych.</small>		
<p>Im Folgenden finden Sie einige Fragen über die Untersuchung selbst und ihr Befinden während der Untersuchung. Bitte beantworten Sie die Formulierungen wahrheitsgemäß.</p>		
<p>Können Sie kurz beschreiben, wozu Sie in dieser Untersuchung aufgefordert wurden?</p> <p>_____</p> <p>_____</p>		
<p>Bitte geben Sie an, inwieweit eine Formulierung auf Sie zutrifft. Hierfür kreuzen Sie bitte auf dem Zahlenstrahl den entsprechenden Wert von Null (0= minimal, im Sinne von gar nicht) bis Hundert (100= maximal, im Sinne von dem stärksten Ausprägungsgrad) an.</p>		
Während der Untersuchung...	0 _____ 100	
... war ich müde.	_____	
... war ich gelangweilt.	_____	
... hatte ich Interesse an den Tests.	_____	
... stand ich der Untersuchung misstrauisch gegenüber.	_____	
... habe ich versucht die Aufgaben so gut wie möglich wie ein Patient mit einer kognitiven Störung zu bearbeiten.	_____	
Während der Untersuchung...	0 _____ 100	
.... habe ich versucht zu simulieren.	_____	
... habe ich die Aufgaben schlechter bearbeitet als ich es gekonnt hätte.	_____	
... war ich langsamer als ich es gekonnt hätte.	_____	
... habe ich weniger Wörter erinnert als ich eigentlich gewusst hätte.	_____	