

Aus der Klinik für Psychiatrie und Psychotherapie  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

**Der Berliner Shape Trail Test – Charakterisierung und  
Validierung eines neuen Trail Making Tests**

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Thomas Günther Riemer

aus Potsdam

Datum der Promotion: 04.09.2015

## Inhaltsverzeichnis

<b>Inhaltsverzeichnis.....</b>	<b>2</b>
<b>Abstrakt.....</b>	<b>5</b>
<b>Abstract.....</b>	<b>6</b>
<b>1. Einleitung.....</b>	<b>7</b>
1.1. Allgemeine Einführung.....	7
1.2. Aufbau und Leistungsindikatoren des Trail Making Tests.....	8
1.3. Mit dem Trail Making Test erfasste kognitive Konstrukte.....	9
1.4. Einfluss demografischer Faktoren auf die Leistung im Trail Making Test.....	10
1.5. Normierung des Trail Making Tests.....	11
1.6. Klinische Validierung des Trail Making Tests.....	12
1.6.1. Demenz bei Alzheimerkrankheit.....	12
1.6.2. Trail Making Test bei der Demenz bei Alzheimerkrankheit.....	13
1.6.3. Leichte kognitive Störung.....	14
1.6.4. Trail Making Test bei der leichten kognitiven Störung.....	14
1.7. Abhängigkeit des Trail Making Tests B (TMT-B) von der Kenntnis des Schriftsystems...	15
1.8. Verfahren zum Einsatz unabhängig vom Schriftsystem.....	16
1.8.1. Color Trails Test.....	16
1.8.2. Shape Trail Test.....	19
1.8.3. Trail Making Test Black & White.....	20
1.8.4. Symbol Trail Making Test.....	20
1.8.5. Trail Making Test B, Geometric Version.....	21
1.8.6. Weitere Varianten.....	22
1.9. Zusammenfassung und Ziele der vorliegenden Arbeit.....	22
<b>2. Methodik.....</b>	<b>25</b>
2.1. Untersuchungskollektiv.....	25
2.1.1. Stichprobe der gesunden Probanden.....	25
2.1.2. Stichprobe in der klinischen Studie.....	26
2.2. Untersuchungsinstrumente.....	27
2.2.1. Berliner Shape Trail Test (B-STT).....	28
2.2.2. Neuropsychologische Batterie CERAD-Plus.....	30
2.3. Statistische Methoden.....	32
2.3.1. Auswahl der statistischen Verfahren.....	32

2.3.2. Festlegung des Signifikanzniveaus und Korrektur des Alpha-Fehlers .....	33
2.3.3. Charakterisierung der Stichproben .....	34
2.3.4. Einfluss demografischer Faktoren auf die Bearbeitungszeit im B-STT .....	35
2.3.5. Test-Retest-Reliabilität des B-STT .....	35
2.3.6. Validität .....	35
2.3.7. Unterschiede in den Bearbeitungszeiten von B-STT und TMT-B.....	37
<b>3. Ergebnisse .....</b>	<b>38</b>
3.1. Deskriptive Darstellung der Untersuchungsergebnisse.....	38
3.1.1. Demografische Eigenschaften der Untersuchungsgruppen.....	38
3.1.2. Testleistungen in der Stichprobe der gesunden Probanden.....	40
3.1.3. Testleistungen in der klinischen Studie.....	41
3.1.4. Neuropsychologische Charakteristika in der klinischen Studie.....	41
3.2. Einfluss demografischer Faktoren auf die Bearbeitungszeit im B-STT .....	42
3.2.1. Einfluss von Alter und Bildung (korrelative Untersuchungen) .....	42
3.2.2. Einfluss von Alter und Bildung (multiple Regression).....	43
3.2.3. Einfluss des Geschlechts .....	44
3.3. Test-Retest-Reliabilität des B-STT .....	44
3.4. Validität des B-STT .....	45
3.4.1. Konstruktvalidität.....	45
3.4.2. Kriteriumsvalidität .....	48
3.5. Unterschiede in den Bearbeitungszeiten von B-STT und TMT-B.....	51
3.5.1. Gruppenübergreifende Prüfung.....	51
3.5.2. Vergleich der Untersuchungsgruppen .....	51
<b>4. Diskussion .....</b>	<b>53</b>
4.1. Ziel der Untersuchung .....	53
4.2. Zusammenfassung der Ergebnisse .....	53
4.2.1. Zusammenhang zwischen den demografischen Merkmalen der Probanden und der Bearbeitungszeit im B-STT.....	53
4.2.2. Test-Retest-Reliabilität des B-STT .....	54
4.2.3. Konstruktvalidität des B-STT .....	54
4.2.4. Kriteriumsvalidität des B-STT .....	55
4.2.5. Vergleich der Bearbeitungsdauer von B-STT und TMT-B .....	55
4.3. Kritische Betrachtung der Methoden .....	56
4.3.1. Zusammensetzung und Repräsentativität der Stichprobe der gesunden Probanden .....	56
4.3.2. Zusammensetzung und Repräsentativität der klinischen Stichprobe .....	58

4.3.3. Durchführung der Untersuchung.....	58
4.4. Der B-STT als Instrument in der neuropsychologischen Diagnostik .....	59
4.4.1. Hauptgütekriterien.....	59
4.4.2. Nebengütekriterien .....	61
4.5. Der B-STT als Alternative zum TMT-B .....	64
4.6. Der B-STT im Vergleich zu den bereits veröffentlichten Alternativen zum TMT-B.....	65
4.6.1. Color Trails Test.....	65
4.6.2. Shape Trail Test .....	66
4.6.3. Trail Making Test Black & White.....	66
4.7. Zusammenfassung und Ausblick .....	68
<b>5. Anhang .....</b>	<b>70</b>
5.1. Arbeitsblatt Berliner Shape Trail Test .....	70
5.2. Testinstruktionen.....	71
5.2.1. Trail Making Test A (TMT-A).....	71
5.2.2. Trail Making Test B (TMT-B).....	71
5.2.3. Berliner Shape Trail Test (B-STT).....	71
5.2.4. Instruktionen beim Auftreten von Fehlern .....	72
5.3. Altersunterschiede innerhalb der gesunden Probanden .....	73
5.4. Neuropsychologische Charakteristika in der klinischen Studie – Prüfung der Voraussetzungen für die ANOVA .....	73
5.5. Neuropsychologische Charakteristika in der klinischen Studie – Gruppenunterschiede.....	74
<b>Literaturverzeichnis.....</b>	<b>75</b>
<b>Eidesstattliche Versicherung.....</b>	<b>84</b>
<b>Lebenslauf.....</b>	<b>85</b>
<b>Publikationen.....</b>	<b>86</b>
<b>Danksagung.....</b>	<b>87</b>

## Abstrakt

**Einleitung.** Der Trail Making Test (TMT) ist eines der am häufigsten eingesetzten neuropsychologischen Instrumente in der klinischen Diagnostik kognitiver Störungen. Die Aufgabe besteht im Verbinden einer Zahlenfolge (TMT-A) bzw. im alternierenden Verbinden einer Zahlen- und einer Buchstabenfolge (TMT-B). Bedingt durch die alphabetische Folge beim TMT-B ergeben sich Limitationen hinsichtlich der Anwendbarkeit des Tests im interkulturellen Kontext sowie bei Analphabetismus. In der vorliegenden Arbeit wurde mit dem Berliner Shape Trail Test (B-STT) eine Variante des Trail Making Tests validiert, die auf Buchstaben als Stimulusmaterial verzichtet. Ziel der Studie war es, die Einflüsse demografischer Faktoren auf die Bearbeitungszeit im B-STT zu untersuchen und Reliabilität, Validität und Ökonomie des Verfahrens zu überprüfen.

**Methodik.** Für die Studie wurden 235 gesunde Probanden im Alter von 20–80 Jahren und 118 Patienten im Alter von 65–80 Jahren, darunter 82 Patienten mit leichter kognitiver Störung (MCI) und 36 Patienten mit Demenz bei Alzheimerkrankheit (DAT), rekrutiert. Der B-STT wurde als Teil einer neuropsychologischen Testbatterie eingesetzt, die in jedem Falle den TMT-A und TMT-B sowie subgruppenabhängig weitere Testverfahren enthielt.

**Ergebnisse.** Bei den gesunden Probanden bestanden signifikante Zusammenhänge zwischen der Bearbeitungszeit im B-STT und dem Alter und Bildungsniveau, wobei eine langsamere Bearbeitung bei höherem Alter und geringerer Bildung beobachtet wurde. Die Test-Retest-Reliabilität war angemessen. Stichprobenübergreifend zeigten sich hohe Korrelationen zwischen den Bearbeitungszeiten im B-STT und im TMT-B, wobei der B-STT im Vergleich zum TMT-B schneller bearbeitet wurde. Die Bearbeitungszeiten von Patienten mit MCI, Patienten mit DAT und gesunden Probanden im gleichen Altersbereich unterschieden sich signifikant. Dabei zeigten die gesunden Probanden die schnellste und Patienten mit DAT die langsamste Bearbeitung. Analysen der Effektstärken bzw. Trennschärfe indizieren eine gute Diskriminierbarkeit der Untersuchungsgruppen.

**Schlussfolgerung.** Der B-STT erweist sich als geeignetes Instrument in der neuropsychologischen Leistungsdiagnostik und darüber hinaus als valide Alternative zum TMT-B. Durch die schnellere Bearbeitung besteht zudem ein Vorteil hinsichtlich der Testökonomie. Vor dem klinischen Einsatz in der Diagnostik kognitiver Störungen ist jedoch eine Normierung des B-STT erforderlich.

## Abstract

**Introduction.** The Trail Making Test (TMT) is one of the most frequently used neuropsychological tasks for the clinical assessment of cognitive function. In the TMT, subjects are required to connect numbers (TMT-A) or to alternately connect numbers and letters (TMT-B) in sequence. Owing to its reliance on the alphabetic sequence, the TMT-B is of limited use in a cross-cultural context and in the examination of patients with analphabetism. In the present study, the Berlin Shape Trail Test (B-STT), a new variant of the TMT which does not use letters as stimuli, is validated. The study aims to characterise the demographic influences on the completion time of the B-STT and to explore the reliability, validity, and test efficiency.

**Methods.** 235 healthy subjects aged 20–80 years and 118 patients aged 65–80 years participated in the study. The patient group consisted of two subpopulations, 82 patients with mild cognitive impairment (MCI) and 36 patients with Alzheimer-type dementia (DAT). The B-STT was employed as part of a neuropsychological assessment battery, which consisted of TMT-A and TMT-B and, depending on the subpopulation, further cognitive testing procedures.

**Results.** In healthy subjects, age and education were significantly correlated with completion time in the B-STT, with slower performance in older and less educated subjects. Test-retest reliability was adequate. In all groups, completion times of B-STT and TMT-B were highly correlated, while the B-STT was completed faster than the TMT-B. Completion times differed significantly between patients with MCI, patients with DAT, and healthy subjects of the same age. Fastest completion times were observed in healthy subjects, while patients with DAT required the most time. Effect size and power analyses indicated a good distinguishability between the groups.

**Conclusion.** The B-STT has proved to be an adequate instrument in neuropsychological assessment and, beyond that, a valid alternative to the TMT-B. In addition, it has a distinct advantage with regard to test efficiency compared to the TMT-B. Prior to systematic application in clinical settings, however, normative data for the B-STT has to be collected.

# 1. Einleitung

## 1.1. Allgemeine Einführung

In den letzten Jahrzehnten lässt sich in Deutschland eine zunehmende Alterung der Bevölkerung beobachten. Erwartet wird, dass sich dieses Phänomen auch in Zukunft fortsetzen wird. So wird prognostiziert, dass im Jahr 2050 ein Drittel der Bevölkerung 60 Jahre oder älter sein wird [1]. Diese Entwicklung ist dabei nicht spezifisch für Deutschland, sondern tritt in vielen Industrienationen auf [2]. Dieser sogenannte demografische Wandel führt auch zu einer erhöhten Prävalenz altersassoziierter Erkrankungen. Darunter sind dementielle Erkrankungen von besonderer Bedeutung: Während die Häufigkeit dementieller Erkrankungen bei 65- bis 69-Jährigen etwa 1,5 % beträgt, so verdoppelt sie sich jeweils im Intervall von fünf Altersjahren und erreicht bei 90-Jährigen einen Wert von ca. 30 % [3]. Eine Schätzung des Robert Koch-Instituts ging von einer Prävalenz von ca. 900.000 Demenzkranken und einer Inzidenz von ca. 200.000 Fällen bereits im Jahr 2002 aus [4]. Die hohe Zahl betroffener Patienten verdeutlicht die Notwendigkeit, Demenzerkrankungen sicher zu identifizieren, um gegebenenfalls eine spezifische Therapie einleiten zu können.

In der Diagnostik dementieller Syndrome hat die neuropsychologische Leistungsdiagnostik neben klinischen und paraklinischen Untersuchungen eine besondere Relevanz [5]. Der Trail Making Test (TMT) ist eines der international am häufigsten eingesetzten neuropsychologischen Instrumente zur Erfassung von kognitiven Störungen [6,7]. Der Name „Trail Making Test“ wurde erstmals 1944 bei seiner Veröffentlichung in der *Army Individual Test Battery* verwendet; Prototypen des Verfahrens existierten jedoch bereits zuvor unter den Namen *Taylor Number Series*, *Test of Distributed Attention* und *Partington Pathway Test* [8,9]. Heutzutage wird der TMT einzeln als Screening-Test oder gemeinsam mit anderen Verfahren als Teil einer neuropsychologischen Testbatterie eingesetzt [10]. So ist er beispielsweise Bestandteil der erweiterten neuropsychologischen Batterie des *Consortium to Establish a Registry for Alzheimer's Disease* (CERAD-Plus; [11]) oder der Erweiterung der kognitiven Skala der *Alzheimer's Disease Assessment Scale* (ADAS-Cog-Plus; [12]), zweier Verfahren, die für die Diagnose und Schweregradbestimmung dementieller Syndrome eingesetzt werden.

## ***1.2. Aufbau und Leistungsindikatoren des Trail Making Tests***

Der TMT setzt sich aus zwei Untertests (TMT-A, TMT-B) zusammen, die den Probanden auf separaten Arbeitsblättern (DIN A4-Format, 210 mm × 297 mm) vorgelegt werden. Beim TMT-A befinden sich auf dem Arbeitsblatt 25 Kreise, die jeweils eine der Zahlen 1–25 enthalten. Die Aufgabe besteht darin, die Kreise so schnell wie möglich in aufsteigender Reihenfolge zu verbinden (1–2–3–4–...). Beim TMT-B befinden sich auf dem Arbeitsblatt ebenfalls 25 Kreise, die jedoch die Zahlen 1–13 und die Buchstaben A–L enthalten. Die Aufgabe des Probanden ist es, die Kreise so zu verbinden, dass sich die Zahlen und Buchstaben abwechseln und dabei die numerische bzw. alphabetische Folge eingehalten wird (1–A–2–B–...). Die Testformen werden jeweils zunächst anhand einer Übungsversion mit acht Kreisen erläutert. Bei der Bearbeitung der eigentlichen Arbeitsblätter wird die Bearbeitungszeit gemessen. Unterlaufen den Probanden Fehler, so werden sie durch den Testleiter darauf hingewiesen, ohne dass die Lösung offenbart wird [6]. Teilweise wird der TMT nach Erreichen einer bestimmten Bearbeitungsdauer abgebrochen. So wird im Manual der CERAD-Plus eine maximale Bearbeitungszeit von drei Minuten für den TMT-A und fünf Minuten für den TMT-B angegeben, wobei die Zeitmessung unmittelbar mit dem Vorlegen des jeweiligen Arbeitsblatts und der Demonstration des ersten Items beginnt [11].

Bei der Durchführung des TMT werden für jede der beiden Testformen Bearbeitungszeit und Fehlerzahl ermittelt. Die Bearbeitungszeiten sind als Leistungsindikatoren relevant. Zusätzlich können daraus weitere Parameter ermittelt werden: Sowohl der Differenzwert B–A als auch der Quotient B/A zielen darauf, die Leistung des TMT-B um diejenigen Aspekte zu bereinigen, die in TMT-A und TMT-B gleichermaßen erfasst werden. Dadurch spiegelt der resultierende Indikator vornehmlich die zusätzlichen Anforderungen des TMT-B wider [13]. Seltener wird weiterhin der Proportionalwert  $(B-A)/A$  verwendet, der als sensibles Maß für die Integrität des präfrontalen Kortex beschrieben wird [14]. Hinsichtlich der Fehler wird in der Literatur über drei unterschiedliche Typen berichtet, nämlich (1) Proximitätsfehler, d. h. Fortsetzung der Folge mit einem räumlich nahegelegenen Zeichen im TMT-A oder TMT-B, (2) Perseverationsfehler, d. h. Fehler beim Wechsel zwischen Zahl und Buchstabe im TMT-B, und (3) Sequenzfehler, d. h. Fortsetzung der Folge unter Auslassung von Zeichen in TMT-A oder TMT-B [15]. Auf eine spezifische Betrachtung der Fehler wird jedoch im klinischen Kontext oft verzichtet. Die Fehlerzahl geht vielmehr in die Bearbeitungszeit ein, da auf Fehler in der Standardanweisung

durch den Untersucher hingewiesen wird und sich aufgrund der Korrektur durch den Probanden eine Verlängerung der Bearbeitungszeit ergibt [15].

### ***1.3. Mit dem Trail Making Test erfasste kognitive Konstrukte***

Die Leistung im TMT wird mit verschiedenen kognitiven Domänen in Verbindung gebracht, die meist mithilfe korrelativer Ansätze, Regressions- oder Faktorenanalysen ermittelt wurden [16]. Beiden Testformen liegen zunächst basale visuelle Wahrnehmungs- und Suchprozesse sowie feinmotorische Fähigkeiten [17,18] zugrunde. Die damit offensichtliche Abhängigkeit der Testleistungen von der Intaktheit des motorischen und des visuellen Systems kann problematisch sein: Alters- oder krankheitsbedingte Einbußen in diesen Bereichen können zur Verlängerung der Bearbeitungszeiten beitragen und die Interpretierbarkeit des Testergebnisses erschweren [19]. So zeigen sich negative Assoziationen zwischen den Testleistungen und Parametern der allgemeinen körperlichen Leistungsfähigkeit bzw. Mobilität wie Handgriffstärke und maximaler Gehgeschwindigkeit [20].

TMT-A und TMT-B erfassen zudem allgemeine Intelligenz [21] und Aufmerksamkeit bzw. kognitives Tempo [18,22]. Darüber hinaus stellt der TMT-B zusätzliche Anforderungen an exekutive Funktionen [10,18,23]. Bezüglich der genauen exekutiven Komponenten besteht kein Konsens: In verschiedenen Untersuchungen wurden als zusätzliche Anforderungen Arbeitsgedächtnis [16,24], kognitive Flexibilität [24,25] sowie Inhibition und Aufmerksamkeitswechsel [26,27] genannt. Die ermittelten Komponenten variieren dabei aufgrund unterschiedlicher in die Analysen eingeschlossener Bezugsvariablen [28].

Bei der Bildung des Differenzwertes  $B-A$  und des Quotienten  $B/A$  wird davon ausgegangen, dass die beiden Testformen mit Ausnahme des im TMT-B notwendigen Wechsels zwischen der numerischen und der alphabetischen Folge äquivalent sind. Dieser Annahme widersprechen Untersuchungen, bei denen die räumlichen Stimuluskonfigurationen von TMT-A und TMT-B vertauscht wurden [29,30]. Hierbei wird offenkundig, dass die im Mittel längere Bearbeitungszeit im TMT-B teilweise auf seine größere räumliche Komplexität zurückgeführt werden kann. So sind der mittlere Abstand zwischen aufeinanderfolgenden Stimuli, die Gesamtlänge nach Verbindung aller Stimuli sowie die Zahl im Umkreis eines Stimulus liegender Distraktoren jeweils im TMT-B höher als im TMT-A [29]. Die erhöhten Anforderungen des TMT-B resultieren dementsprechend nicht nur aus der zusätzlichen Beteiligung exekutiver Funktionen.

#### ***1.4. Einfluss demografischer Faktoren auf die Leistung im Trail Making Test***

Zahlreiche Untersuchungen widmen sich der Fragestellung, inwieweit sich Merkmale der Testperson wie Alter, Bildungsniveau, Geschlecht und Ethnizität auf die Leistungen im Trail Making Test auswirken. Im *Handbook of Normative Data for Neuropsychological Assessment, 2nd edition* [23], welches eine Sammlung von Normierungsstudien zum Trail Making Test aus den Jahren 1968 bis 2004 enthält, wird subsummiert, dass sich Alter und Bildungsniveau, nicht aber das Geschlecht der Probanden auf ihre Leistungen im Trail Making Test auswirken. Ein höheres Alter und eine kürzere Bildungsdauer sind dabei mit geringeren Testleistungen assoziiert. Diese Einschätzung bestätigt sich in aktuelleren, in verschiedenen Ländern durchgeführten Untersuchungen, die eine Zunahme der Bearbeitungszeiten in TMT-A und TMT-B sowie eine erhöhte Fehleranfälligkeit im TMT-B mit höherem Alter und geringerer Bildung, jedoch einen allenfalls geringfügigen und in seiner Richtung nicht konsistenten Einfluss des Geschlechts der Testperson zeigen [15,24,31–34].

Bezüglich des Einflusses des ethnisch-kulturellen Hintergrundes liegen relativ wenige Studien vor, die zudem teilweise an klinischen Populationen erhoben wurden. Insgesamt ergeben sich Hinweise auf eine schnellere Bearbeitung des TMT-A [35] bzw. des TMT-A und des TMT-B [36,37] bei US-Amerikanern europäischer Abstammung (*Caucasians*) im Vergleich zu Afroamerikanern. Im Vergleich zu US-Amerikanern hispanoamerikanischer oder spanischer Abstammung (*Hispanics*) lassen sich bei *Caucasians* teils vergleichbare [35], teils schnellere Bearbeitungszeiten in TMT-A und TMT-B [37,38] nachweisen. Im Gegensatz dazu zeigen sich keine Unterschiede zwischen *Hispanics*, Afroamerikanern und Amerikanern asiatischer Abstammung [35,39] und zwischen verschiedenen Populationen mit spanischer Muttersprache [40]. Die dargestellten Befunde weisen insgesamt darauf hin, dass die an einer bestimmten ethnischen Gruppe erhobenen Normen nur mit Vorsicht für die Interpretation von Leistungen einer Person anderer Ethnizität oder Kultur verwendet werden dürfen. Zu diesem Schluss kommen auch die Autoren einer 2008 veröffentlichten Untersuchung zum Vergleich von Normdaten des Trail Making Tests in zehn verschiedenen Ländern: Auch nach Alters- und Bildungsadjustierung zeigten sich erhebliche Unterschiede zwischen verschiedenen Populationen [41].

### ***1.5. Normierung des Trail Making Tests***

Die in Abschnitt 1.4. dargestellten Befunde zeigen, dass die Leistung im Trail Making Test vorrangig durch Alter und Bildungsniveau der Testperson, populationsabhängig aber auch durch den ethnischen Hintergrund beeinflusst wird. Bei der Erhebung von Normdaten ist daher eine explizite Berücksichtigung dieser Faktoren indiziert, auch wenn teilweise diskutiert wird, ob ein hohes Maß an Differenzierung eine Verbesserung der Genauigkeit nur vortäuscht [6]. In der Praxis erfolgt eine Stratifizierung in der Regel mindestens nach Altersgruppen, teilweise auch unter Berücksichtigung weiterer Merkmale. An dieser Stelle werden die beiden im deutschen Sprachraum zur Verfügung stehenden Normdatensammlungen vorgestellt:

Im Rahmen der Normierung der erweiterten neuropsychologischen Batterie des *Consortium to Establish a Registry for Alzheimer's Disease* (CERAD-Plus) wurden durch die Arbeitsgruppe der Memory Clinic Basel Normdaten an 604 gesunden, deutschsprachigen Probanden (270 Frauen, Alter 50–88 Jahre) erhoben [11]. Neben einer tabellarischen Übersicht der Normwerte, die nach drei Altersbereichen, zwei Bildungsstufen sowie dem Geschlecht stratifiziert ist, ist nach Registrierung ([www.memoryclinic.ch](http://www.memoryclinic.ch)) eine differenzierte Auswertung mithilfe eines Regressionsmodells verfügbar.

Eine zweite Normierungsstudie, die im Unterschied zur Baseler Normierung einen größeren Altersbereich Erwachsener abdeckt, wurde 2012 veröffentlicht [24]. Die verwendete Stichprobe setzt sich aus 405 gesunden, deutschsprachigen Probanden (223 Frauen, Alter 18–84 Jahre) zusammen. Normwerte werden nach vier Altersbereichen und zwei Bildungsstufen vergeben. Eine zusätzliche Berücksichtigung des Geschlechts erfolgt nicht, da in der Untersuchung kein Effekt des Geschlechts nachgewiesen wurde. Relevant ist, dass die Autoren zwar die Standardversion des TMT-A verwendeten, jedoch eine modifizierte Version des TMT-B, bei der die Gesamtstreckenlänge und die durchschnittliche Entfernung zwischen zwei benachbarten Items dem TMT-A angeglichen wurde. Dementsprechend können die Normwerte des modifizierten TMT-B nicht für die Standardform des Verfahrens verwendet werden.

Die genannten Normierungen beziehen sich ausschließlich auf die Bearbeitungszeiten. Die Fehleranfälligkeit geht also nur indirekt, d. h. über eine Verlängerung der Bearbeitungszeit, in die Testleistung ein.

## ***1.6. Klinische Validierung des Trail Making Tests***

Seitdem die Utilität des Trail Making Tests in der Diagnostik zerebraler Funktionsstörungen belegt wurde [42], wird das Verfahren bei vielfältigen neurologischen und psychiatrischen, aber auch primär internistischen Störungsbildern eingesetzt. In den folgenden Abschnitten werden die Befunde zu Veränderungen der Leistungen im Trail Making Test bei Vorliegen einer Demenz bei Alzheimerkrankheit bzw. einer leichten kognitiven Störung zusammengefasst. Beide Erkrankungen werden kurz vorgestellt, wobei hier auf eine umfassende Abhandlung der Störungsbilder verzichtet wird. Problematisch ist, dass der TMT selbst als neuropsychologisches Instrument zur Diagnosestellung eingesetzt wird, wodurch sich prinzipiell die Gefahr einer Zirkularität zwischen Defizitbefund und Diagnose ergibt. Es werden daher, sofern möglich, vorrangig die Untersuchungen dargestellt, bei denen der TMT nicht selbst als Instrument zur Diagnosefindung verwendet wurde.

### ***1.6.1. Demenz bei Alzheimerkrankheit***

Demnzerkrankungen werden in der *International Statistical Classification of Diseases and Related Health Problems, 10th revision* (ICD-10) allgemein als Folge einer meist chronischen oder progredienten Schädigung des Gehirns mit Beeinträchtigung vieler kognitiver Funktionen und gegebenenfalls Veränderungen der emotionalen Kontrolle, des Sozialverhaltens oder der Motivation charakterisiert [43]. Mit einem Anteil von etwa 60 % ist die Demenz bei Alzheimerkrankheit (Demenz vom Alzheimerstyp; DAT) die häufigste Form der Demnzerkrankungen [44,45]. Es handelt sich um eine schleichend progrediente neurodegenerative Erkrankung, die auf mikroskopischer Ebene durch intrazelluläre Bündelungen von hyperphosphoryliertem Tau-Protein (neurofibrilläre Tangles) sowie ein deutlich vermehrtes Auftreten extrazellulärer Aggregationen von Beta-Amyloid (senile Plaques) gekennzeichnet ist [45]. Makroskopisch lässt sich eine Hirnvolumenabnahme beobachten, die sich ausgehend vom mesiotemporalen Kortex einschließlich der Hippokampusformation über die gesamte Hirnoberfläche ausdehnen kann [45]. Indikatoren für das Vorliegen einer DAT bereits im frühen Stadium sind ein verringerter Metabolismus im temporoparietalen Kortex sowie verringerte Konzentrationen von Amyloid  $\beta_{1-42}$  und erhöhte Konzentrationen an Tau-Protein und Phospho-Tau im Liquor cerebrospinalis [46]. Neuropsychologisch imponieren schon im frühen Krankheitsstadium neben mnestischen Defiziten auch Einbußen einiger Facetten der sprachlichen sowie Aufmerksamkeits- und Exekutivfunktionen; späterhin kommt es neben einer Zunahme dieser Defizite zu Störungen der Raumverarbeitung, Gnosie und Praxis [47–49].

Nach der *National Institute of Neurological and Communicative Disorders and Stroke – The Alzheimer's Disease and Related Disorders Association* (NINCDS-ADRDA) kann die sichere Diagnose einer Demenz bei Alzheimerkrankheit erst durch eine postmortale neuropathologische Untersuchung gestellt werden. Notwendige Voraussetzung für die Diagnose einer klinisch wahrscheinlichen Demenz bei Alzheimerkrankheit sind (1) das Vorliegen eines dementiellen Syndroms in der klinischen und neuropsychologischen Untersuchung, (2) Defizite in mindestens zwei kognitiven Bereichen, (3) eine fortschreitende Verschlechterung des Gedächtnisses und anderer kognitiver Funktionen mit (4) Beginn im 40.–90. Lebensjahr, (5) das Fehlen einer Bewusstseinsbeeinträchtigung und (6) der Ausschluss einer anderen die Symptomatik erklärenden somatischen Erkrankung [50].

### *1.6.2. Trail Making Test bei der Demenz bei Alzheimerkrankheit*

Unter verschiedenen Instrumenten zur Erfassung exekutiver Funktionen ist der Trail Making Test besonders sensitiv bei der Detektion von dysexekutiven Störungen bei Patienten mit beginnender DAT [51]. Dadurch erlangt der TMT als diagnostisches Instrument zum Nachweis von kognitiven Beeinträchtigungen bei DAT besondere Relevanz und zählt zu den wichtigsten und am häufigsten eingesetzten Instrumenten bei der Diagnosefindung [52,53].

Tatsächlich zeigen sich im TMT-B und auch im TMT-A besonders ausgeprägte Leistungseinbußen bereits in präklinischen und frühen Stadien der DAT [54]. Bei betroffenen Patienten lassen sich – verglichen mit gesunden Probanden, aber auch gegenüber Patienten mit leichter kognitiver Störung, – verlängerte Bearbeitungszeiten in TMT-A und TMT-B nachweisen [15,55,56]. Im Vergleich zu gesunden Probanden zeigt sich zudem eine erhöhte Fehleranfälligkeit im TMT-B [15,55]. Weiterhin ist belegt, dass das Ausmaß der bei DAT nachweisbaren Leistungseinbußen im TMT mit der Konfiguration bestimmter Biomarker im Liquor cerebrospinalis sowie mit Funktionseinbußen spezifischer Hirnareale in Verbindung steht: Stärkere Minderleistungen im TMT korrespondieren dabei mit einem niedrigeren Gehalt an Amyloid  $\beta_{1-42}$  und einem höheren Gehalt an Tau-Protein und Phospho-Tau im Liquor [57] sowie Minderperfusionen in bestimmten Hirnarealen (TMT-A: Lobulus parietalis superior [58]; TMT-B: anteriorer cingulärer Cortex, Nucleus caudatus, Thalamus [59]).

### *1.6.3. Leichte kognitive Störung*

Die leichte kognitive Störung (mild cognitive impairment; MCI) ist ebenfalls durch von der Norm abweichende Einbußen kognitiver Funktionen gekennzeichnet, wobei diese nicht das Ausmaß eines dementiellen Syndroms erreichen und zu keiner wesentlichen Beeinträchtigung der Alltagskompetenz führen [43,60]. Die Diagnose sollte nur in Verbindung mit einer organischen Grunderkrankung und nicht bei Vorliegen einer psychiatrischen Diagnose aus dem Abschnitt F10–F99 gestellt werden [43]. Je nach Art der neuropsychologischen Defizite wird zwischen amnestischer und nicht-amnestischer Form unterschieden. Zusätzlich wird nach der Anzahl der betroffenen neuropsychologischen Bereiche zwischen MCI mit Beteiligung einer kognitiven Domäne bzw. multipler Domänen unterschieden [61]. Der Zustand einer MCI wird von einigen Autoren als Übergangsstadium zwischen altersentsprechender Hirnleistung und beginnender DAT angesehen [62] und stellt in jedem Fall ein erhöhtes Risiko für die Entwicklung einer DAT dar [60].

### *1.6.4. Trail Making Test bei der leichten kognitiven Störung*

Patienten mit MCI zeigen im Vergleich zu gesunden Probanden ein geringeres Bearbeitungstempo sowohl im TMT-A als auch im TMT-B; im TMT-B kommt es zusätzlich zu einer erhöhten Fehlerrate [15,61,63]. Das Ausmaß der Einbußen ist dabei subtypenabhängig: Bereits beim amnestischen Subtyp der MCI ist eine Verlängerung der Bearbeitungszeiten im TMT-B [64,65] bzw. in TMT-A und TMT-B [56,66] nachweisbar. Noch deutlichere Einbußen als beim amnestischen Subtyp finden sich jedoch in beiden Testformen bei Patienten mit dem dysexekutiven Subtyp der nicht-amnestischen Form der MCI [67] sowie bei Vorliegen von Einbußen in multiplen Domänen [68].

Von besonderem Interesse ist die Frage, ob Leistungsunterschiede bei Patienten mit MCI in verschiedenen neuropsychologischen Verfahren mit unterschiedlichen Risiken für die Entstehung eines Demenzsyndroms korrespondieren. In einer aktuellen Längsschnittstudie wurde das Konversionsrisiko für Patienten mit MCI vom amnestischen Subtyp zur DAT unter Beachtung verschiedener Biomarker und neuropsychologischer Parameter untersucht [69]. Dabei zeigte sich, dass der TMT-B unter den aufgenommenen neuropsychologischen Verfahren den höchsten prädiktiven Wert für die Entwicklung einer DAT aufwies. Gleichsinnig zeigten sich in zwei Längsschnittuntersuchungen über 24 Monate signifikant längere Bearbeitungszeiten im

TMT-B [70] bzw. TMT-A und TMT-B [71] bei Patienten, die von einer MCI zu einer DAT konvertierten, im Vergleich zu Patienten mit unverändert bestehender MCI.

### ***1.7. Abhängigkeit des Trail Making Tests B (TMT-B) von der Kenntnis des Schriftsystems***

Die weite Verbreitung und der beschriebene regelmäßige Einsatz des Trail Making Tests in der neuropsychologischen Leistungsdiagnostik erfordern eine kritische Auseinandersetzung mit dem Testverfahren, die die korrekte Erfassung des Leistungsniveaus der Probanden und die Durchführbarkeit bei verschiedenen Populationen betrifft.

Im klinischen Alltag ist die Anwendbarkeit des Trail Making Tests durch die Art des verwendeten Stimulusmaterials limitiert. Dies betrifft insbesondere den TMT-B, bei dem das lateinische Alphabet als überlernte Folge zur Lösung der Aufgabe erforderlich ist. Dies führt zu einer systematischen Benachteiligung von Personen mit Analphabetismus sowie von Personen aus Kulturkreisen mit anderen Schriftsystemen [72]. Folglich ist zumindest der TMT-B bei diesen Personen nicht valide einsetzbar ist.

Am Beispiel Deutschlands sei die Relevanz dieser Einschränkung der Anwendbarkeit verdeutlicht: Nach einer im Auftrag der Bundesregierung veröffentlichten Expertise des Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung lebten im Jahr 2009 in Deutschland etwa 16 Millionen Personen mit Migrationshintergrund. Darunter zählten Zugezogene aus Griechenland, Italien, Kroatien und der Türkei zu der am längsten in Deutschland wohnhaften Personengruppe [2]. Auch wenn keineswegs bei all diesen Personen davon ausgegangen werden kann, dass Lücken bezüglich des lateinischen Alphabets vorliegen, wird klar, dass die potentielle Einsetzbarkeit des TMT-B bedeutsam eingeschränkt ist. Das trifft auch dann zu, wenn Personen im Rahmen eines Fremdsprachenerwerbs das lateinische Alphabet erlernt haben. So zeigte sich in einer Untersuchung bei gesunden englischen Muttersprachlern britischer Herkunft und Chinesen mit englischen Sprachkenntnissen eine geringere Performanz der chinesischen Probanden im TMT-B, jedoch nicht im TMT-A [73]. Die Schriftsystemabhängigkeit stellt damit eine erhebliche Schwierigkeit bei der Anwendung des TMT-B dar. Aufgrund der Relevanz dieses Themas für die vorliegende Arbeit sind mögliche Strategien zur Lösung dieses Problems nachfolgend ausführlicher dargestellt.

## ***1.8. Verfahren zum Einsatz unabhängig vom Schriftsystem***

Dem Problem der Schriftsystemabhängigkeit des Trail Making Tests B wird teilweise durch die Entwicklung von Testformen in weiteren Alphabeten begegnet. So liegen beispielsweise arabische [74], griechische [34], hebräische [75], japanische [76], koreanische [77], tschechische [32] und türkische [78] Versionen vor. Weiterhin existieren chinesische Formen, in denen die Buchstaben des TMT-B durch chinesische Zahlzeichen [79] bzw. chinesische Sternzeichen [80] ersetzt wurden. Wie die ursprüngliche Version des Trail Making Tests beschränkt sich die Einsetzbarkeit dieser Versionen auf ihre linguistische Zielgruppe und auf Personen mit Kenntnis des jeweiligen Alphabets bzw. Zeichensystems.

Eine umfassendere Strategie zur Lösung des Problems liegt in der Entwicklung von Verfahren, die gänzlich auf die Verwendung von Schriftzeichen verzichten. Dadurch wird einerseits die Kulturabhängigkeit der Methode reduziert; andererseits ermöglicht dies den Einsatz auch bei mangelhafter Kenntnis des Alphabets. Im Folgenden werden verschiedene Testverfahren vorgestellt, die mit diesem Grundgedanken entwickelt wurden.

### *1.8.1. Color Trails Test*

Der Color Trails Test (CTT; [81]) wurde ursprünglich im Rahmen einer multizentrischen, kulturübergreifenden Studie zur Erforschung neuropsychologischer Aspekte der HIV-Infektion entwickelt. Hauptziel der Testentwicklung war die Verminderung kultureller Einflussfaktoren. Dies wurde durch den Verzicht auf Schriftzeichen sowie durch die Verwendung fakultativ nonverbal durchführbarer Instruktionen realisiert. Ein weiterer Aspekt bei der Entwicklung dieses Tests bestand in der Einschränkung der Fotokopierbarkeit der Testunterlagen, um einem Qualitätsverlust durch wiederholtes Kopieren entgegenzuwirken [81].

Der Color Trails Test 1 (CTT-1) ähnelt dem TMT-A; im Unterschied zu diesem sind die Kreise jedoch farbig: Ungerade Zahlen befinden sich in magentafarbenen Kreisen, während alle geraden Zahlen in gelben Kreisen lokalisiert sind. Testpersonen werden instruiert, die Zahlen 1–25 so schnell wie möglich in aufsteigender Reihenfolge zu verbinden, auf die verschiedenfarbigen Hintergründe der Kreise wird bei der Instruktion kein Bezug genommen. Der zweite Teil des Color Trails Tests (CTT-2) dient als Äquivalent zum TMT-B. Hier befindet sich jede der Zahlen 1–24 zweifach auf der Testvorlage, jeweils einmal innerhalb eines magentafarbenen und gelben Kreises, die Zahl 25 kommt nur einmal in einem magentafarbenen Kreis vor. Die Anforderung

besteht darin, die Zahlen in aufsteigender Reihenfolge zu verbinden und dabei zwischen den Kreisfarben zu alternieren (Magenta-1–Gelb-2–Magenta-3–Gelb-4–...). Die Zahlen der jeweils anderen Farbe fungieren als Distraktoren, die Zahl der Stimuli ist also im Vergleich zum TMT-B nahezu verdoppelt. Die Instruktionen für den CTT-1 und CTT-2 sind jeweils auch nonverbal durchführbar. Zu diesem Zweck befindet sich auf den Übungsblättern ein Lösungsschlüssel, der die Lösungssequenz bis zur achten Ziffer darstellt.

Im Handbuch wird der Color Trails Test als ein Verfahren beschrieben, welches es ermöglicht, visuelle Aufmerksamkeit, visuomotorische Geschwindigkeit und exekutive Funktionen in Analogie zum Trail Making Test, jedoch unabhängig vom kulturellen Hintergrund des Probanden zu erfassen [72]. Dennoch zeigen Leistungsunterschiede verschiedener Populationen, dass die in einer Stichprobe ermittelten Referenzdaten nicht ohne weiteres zur Einschätzung der Leistungen in Populationen anderer Kulturen herangezogen werden können [82,83].

Unter den als mögliche Alternative zum Trail Making Test veröffentlichten Verfahren liegt zum Color Trails Test der umfangreichste Literaturbestand vor. Im Folgenden werden Untersuchungen zu demografischen Einflussfaktoren, zur konvergenten Validität des CTT-2 mit dem TMT-B sowie zu Befunden bei Patienten mit Demenz bei Alzheimerkrankheit und bei Patienten mit leichter kognitiver Störung vorgestellt.

### *Demografische Effekte*

Wie beim TMT liegt auch beim CTT eine Alters- und Bildungsabhängigkeit der Testleistungen vor, wobei ein niedriges Alter und eine hohe Bildung mit schnellerer Bearbeitung assoziiert sind. Der Einfluss des Geschlechts scheint wie beim TMT vernachlässigbar zu sein [72,84].

### *Klinische Validierung*

Der Color Trails Test wurde als Instrument zur Erfassung von Leistungsunterschieden in verschiedenen klinischen Populationen eingesetzt. Berichtet werden an dieser Stelle nur die Befunde zur Demenz bei Alzheimerkrankheit sowie zur leichten kognitiven Störung. Bei Patienten mit DAT werden im Vergleich mit gesunden Probanden verlängerte Bearbeitungszeiten sowohl im CTT-1 als auch im CTT-2 sichtbar [85,86]. In einer aktuellen Untersuchung bei Patienten mit MCI zeigten sich zwar im Mittel ebenfalls verlängerte Bearbeitungszeiten; der Unterschied zur gesunden Vergleichsgruppe war jedoch nicht signifikant [86].

---

*Äquivalenz zum Trail Making Test*

Auf konzeptioneller Ebene werden die im CTT-2 höhere Zahl an Stimuli sowie das Vorliegen von Distraktoren von einigen Autoren als relevante Unterschiede herausgestellt, da dadurch stärker visuoperzeptive Funktionen und eine Ablenkbarkeit durch Störreize miteingefasst werden [87,88]. Die Frage der Äquivalenz des CTT-2 zum TMT-B wurde an erwachsenen Probanden und Kindern mithilfe von Zusammenhangsanalysen untersucht. Abhängig von den demografischen Charakteristika der Stichproben offenbart sich eine erhebliche Variabilität der ermittelten Zusammenhänge (schwache bis starke Korrelation). Außerdem wird die Bedeutung ähnlicher Ergebnisse in verschiedenen Studien unterschiedlich interpretiert. Die folgende Darstellung beschränkt sich auf die Untersuchungen an erwachsenen Probanden.

Im Rahmen der ersten Validierung des Color Trails Tests [81] zeigten sich in einer Untersuchung an 30 deutschsprachigen Probanden signifikante moderate Zusammenhänge zwischen den Leistungen im CTT-2 und TMT-B ( $r=.50$ ). Die Autoren bewerteten den Zusammenhang als zufriedenstellend und folgerten, dass der CTT-2 ähnliche Funktionen wie der TMT-B erfasst. In einer späteren Studie, die an 64 Probanden mit türkischer Muttersprache und guten englischen Sprachkenntnissen durchgeführt wurde [87], lag die Stärke des Zusammenhangs ebenfalls im mittleren Bereich ( $r=.45$ ). Die Autoren schätzen die Äquivalenz des CTT-2 zur englischen Version des TMT-B dennoch als fraglich ein. Sie begründen dies mit einem signifikanten Unterschied in den Bearbeitungszeiten der Tests.

Eine Alters-, Bildungs- und Sprachabhängigkeit der Assoziation zwischen den Leistungen im CTT-2 und der englischen Version des TMT-B bildete sich in Untersuchungen an bilingualen chinesischen Probanden und monolingualen englischen Probanden aus Hongkong ab [89,90]. Abhängig von demografischen und sprachlichen Charakteristika zeigten sich deutliche Unterschiede hinsichtlich der Zusammenhänge zwischen CTT-2 und TMT-B ( $r=.27$  bis  $r=.72$ ).

Insgesamt verdeutlicht die vorliegende Literatur, dass eine Äquivalenz der Testverfahren CTT-2 und TMT-B nicht uneingeschränkt angenommen werden kann. Dies ist zum einen durch Unterschiede der Tests in den kognitiven Anforderungen bedingt. Zum anderen scheint der Zusammenhang zwischen den Testleistungen von Alter, Bildung und sprachlichem Hintergrund der Testpersonen abhängig zu sein.

### 1.8.2. *Shape Trail Test*

Wie der Trail Making Test besteht auch der Shape Trail Test (STT; [91]) aus zwei Untertests. Teil A (STT-A) entspricht dabei dem TMT-A, wobei die Positionen der Zahlen 1–25 verändert sind. In Teil B (STT-B) liegt jede der Zahlen 1–25 zweifach vor, wobei sich jede Zahl jeweils einmal in einem Quadrat und einmal in einem Kreis befindet. Die Aufgabe der Testperson ist es, die Zahlen so schnell wie möglich in aufsteigender Folge zu verbinden und dabei zwischen Quadrat und Kreis zu alternieren (Quadrat-1–Kreis-2–Quadrat-3–Kreis-4–...). Das bedeutet, dass wie im CTT-2 die restlichen Zahlen als Distraktoren dienen und die Zahl der Stimuli im Vergleich zum TMT-B doppelt so hoch ist.

Der Shape Trail Test wurde ursprünglich als chinesische Überarbeitung des Trail Making Tests veröffentlicht [92,93]. Publikationen zu diesem Verfahren erfolgten zunächst ausschließlich in chinesischer Sprache. Nach Kenntnis des Autors beinhalten diese Veröffentlichungen Validierungen an verschiedenen klinischen Stichproben, jedoch keine Untersuchung der konvergenten Validität mit der Originalversion des TMT bzw. den Versionen mit chinesischen Zahl- oder Sternzeichen.

In der bisher einzigen in englischer Sprache publizierten Studie erfolgte eine Normierung des Verfahrens sowie Untersuchungen zur Eignung des Verfahrens bei der Diagnostik von kognitiven Störungen [91]. Eingeschlossen wurden insgesamt 2.470 Personen, darunter Probanden ohne kognitive Beeinträchtigungen, Patienten mit leichter kognitiver Störung vom amnestischen Typ sowie Patienten mit beginnender Demenz bei Alzheimerkrankheit. Signifikante Unterschiede zwischen allen Gruppen konnten sowohl beim STT-A als auch beim STT-B nachgewiesen werden, wobei gesunde Probanden die höchsten und Patienten mit DAT die geringsten Leistungen zeigten. Weiterhin zeigte sich innerhalb der gesunden Kontrollgruppe eine geringe, aber signifikante Alters- und Bildungsassoziation der Bearbeitungszeiten, wohingegen Geschlechtseffekte lediglich in einer Subpopulation nachgewiesen werden konnten. Eine Teilstichprobe von 100 kognitiv unbeeinträchtigten Probanden und 50 Patienten mit beginnender Demenz bei Alzheimerkrankheit absolvierte zusätzlich den Color Trails Test. Es zeigten sich jeweils mittelstarke Korrelationen zwischen den beiden Formen der Shape Trail und Color Trails Tests. Dabei bestand der höchste Zusammenhang zwischen dem CTT-2 und dem STT-B ( $r=.57$ ), die Leistung im STT-B war jedoch auch mit den Leistungen im CTT-1 und STT-A moderat assoziiert ( $r=.51$  bzw.  $r=.52$ ). Eine Untersuchung der konvergenten Validität zur

chinesischen Version des TMT-B [73,79] erfolgte auch in dieser Studie nicht. Daher kann keine Aussage zur Äquivalenz des Shape Trail Tests zum Trail Making Test getroffen werden.

### *1.8.3. Trail Making Test Black & White*

Der Ansatz des erst vor kurzem veröffentlichten Trail Making Tests Black & White (TMT-B&W; [94]) ähnelt dem der bereits beschriebenen Verfahren CTT und STT. Im Untertest A (TMT-B&W-A) befinden sich die Zahlen 1–25 jeweils in Kreisen, wobei sich ungerade Zahlen in weißen und gerade Zahlen in schwarzen Kreisen befinden. Die Anforderung besteht im Verbinden der Kreise in der numerischen Folge. Im Untertest B (TMT-B&W-B) liegt jede Zahl mit Ausnahme der Zahl 1 doppelt vor, jeweils einmal in einem weißen und einmal in einem schwarzen Kreis. Die Aufgabe ist es, die Zahlen 1–25 zu verbinden, wobei zwischen schwarzen und weißen Kreisen alterniert werden soll. Die überzähligen Reize fungieren als Distraktoren.

In der veröffentlichten Validierungsstudie wurde der TMT-B&W gemeinsam mit dem TMT und einer neuropsychologischen Testbatterie bei 31 gesunden Kontrollpersonen, 55 Patienten mit leichter kognitiver Störung und 52 Patienten mit Demenz bei Alzheimerkrankheit eingesetzt, wobei nur ein Teil der Personen tatsächlich mit beiden Instrumenten untersucht werden konnte. Signifikante Leistungsunterschiede im TMT-B&W-B konnten zwischen allen Untersuchungsgruppen nachgewiesen werden. Die Korrelation zwischen dem TMT-B und dem TMT-B&W-B war stark ausgeprägt ( $r=.68$ ). Die Autoren bezeichnen den TMT-B&W als angemessene Alternative zum TMT.

### *1.8.4. Symbol Trail Making Test*

Die beschriebenen Verfahren Color Trails Test, Shape Trail Test und Trail Making Test Black & White verzichten zwar auf die Verwendung eines Alphabets, die Testleistung erfordert jedoch die Kenntnis des arabischen Ziffernsystems. Der Grundgedanke des Symbol Trail Making Tests (STMT; [95]) ist es, kulturelle Einflüsse weitestmöglich zu reduzieren. Daher wurden bei der Testentwicklung Zeichen verwendet, welche keine Ähnlichkeit mit Ziffern oder Buchstaben aufweisen. Dementsprechend wird im Unterschied zum Trail Making Test sowie den bereits vorgestellten Alternativen nicht auf überlernte Zeichenfolgen zurückgegriffen.

Der STMT besteht aus drei Untertests: Im ersten Teil (STMT-1) befinden sich, verteilt auf einem Blatt Papier, 24 Kreise mit verschiedenen Symbolen. Am Kopf des Blattes befindet sich zusätzlich ein Lösungsschlüssel, in dem acht dieser Symbole aufgereiht sind. Die Aufgabe ist es,

unter den 24 Kreisen die acht, welche im Lösungsschlüssel abgebildet sind, so schnell wie möglich in der korrekten Reihenfolge zu verbinden. Dabei dienen die anderen 16 Symbole als Distraktoren. Im zweiten Teil (STMT-2) befinden sich erneut 24 Kreise auf dem Papier. Jeweils acht der Kreise sind weiß, grau bzw. mit einem gepunkteten Hintergrund gefüllt. In diesen Kreisen befindet sich jeweils eins von insgesamt acht verschiedenen Symbolen, die nicht aus dem ersten Teil des Tests bekannt sind. Jedes Symbol kommt jeweils einmal in einem weißen, grauen und gepunkteten Kreis vor. Der Lösungsschlüssel am Kopf des Blattes beinhaltet zwei Sequenzen, eine dreiteilige Folge für den Wechsel des Kreishintergrunds (von weiß zu grau zu gepunktet) sowie die Zielsequenz der acht Symbole. Die Anforderung besteht darin, die Symbole so schnell wie möglich entsprechend dem Lösungsschlüssel zu verbinden. Dabei werden zunächst die drei Kreise, die das erste Symbol beinhalten, in der vorgegebenen Reihenfolge der Hintergründe (d. h. von weiß zu grau zu gepunktet) verbunden, anschließend erfolgt die Verbindung zum zweiten Symbol mit weißem, grauem und gepunktetem Hintergrund. Dies setzt sich bis zum achten Symbol mit gepunktetem Hintergrund fort. Der dritte Teil des Tests (STMT-3) stellt eine Aufgabe zum impliziten Gedächtnis dar: Die Probanden sollen diejenigen Symbole identifizieren, mit denen sie in den vorherigen Teilen des Tests konfrontiert wurden.

Im Rahmen der bisher einzigen Veröffentlichung zum STMT [95] erfolgte neben einer Normierung und einer Prüfung der Test-Retest-Reliabilität auch eine Validierung am Trail Making Test. Bezüglich der Normierung ist bemerkenswert, dass sich die Testleistungen unabhängig vom sprachlichen Hintergrund sowie Bildungsniveau zeigten, ein Alterseffekt wurde zudem lediglich für die Rekognitionsaufgabe berichtet. Die beim Vergleich mit dem TMT ermittelten Korrelationskoeffizienten sprechen für einen mittleren Zusammenhang zwischen STMT-1 und TMT-A ( $r=.42$ ) bzw. TMT-B ( $r=.45$ ) und einen niedrigen Zusammenhang zwischen STMT-2 und TMT-A ( $r=.19$ ) bzw. TMT-B ( $r=.23$ ). Die Autoren schließen daraus, dass mit den Verfahren unterschiedliche Konstrukte erfasst werden und der STMT nicht als Alternative zum TMT geeignet ist.

#### *1.8.5. Trail Making Test B, Geometric Version*

Nicht spezifisch als schriftunabhängiges Verfahren publiziert, aber konzeptuell einem solchen entsprechend ist der *Trail Making Test B, Geometric Version* (TMT-G; [96]). Bei diesem liegen auf dem Arbeitsblatt die Zahlen 1–8 dreifach vor, dabei jeweils einmal in einem Kreis, einem Quadrat und einem Kreuz. Die Anforderung besteht darin, die Figuren in aufsteigender Reihenfolge so zu verbinden, dass jede Zahl dreimal nacheinander vorkommt und dabei die

Sequenz Kreis–Quadrat–Kreuz eingehalten wird (Kreis-1–Quadrat-1–Kreuz-1–Kreis-2–...). Die Sequenz der Figuren befindet sich an der Oberseite des Arbeitsblattes. In der einzigen zum TMT-G veröffentlichten Untersuchung wurde der Test als zusätzliches Maß nach dem TMT-A und dem TMT-B an insgesamt 300 gesunden Probanden durchgeführt. Neben einer hohen Korrelation zum TMT-B ( $r=.80$ ) berichten die Autoren eine Alters- und Bildungsabhängigkeit der Testleistungen [96]. Das Verfahren wurde nicht an einer klinischen Population validiert.

### *1.8.6. Weitere Varianten*

Gelegentlich werden in einzelnen Studien modifizierte Versionen des TMT-B verwendet, bei denen die Buchstaben durch Kürzel der Monatsnamen (z. B. Jan, Feb, Mar; [97]) bzw. Wochentage (z. B. Mon, Tue, Wed; [98]) ersetzt werden. Für letztere Form muss der Test zusätzlich in seiner Gesamtlänge reduziert werden. Diese Varianten eignen sich prinzipiell, wenn die Probanden zwar lesen können, aber die alphabetische Folge nicht beherrschen. Fraglich ist, inwieweit bei Übersetzung des Verfahrens in verschiedene Sprachen Kürzel äquivalenter Länge und Diskriminierbarkeit zur Verfügung stehen.

## **1.9. Zusammenfassung und Ziele der vorliegenden Arbeit**

Der Trail Making Test ist eines der weltweit am häufigsten eingesetzten Instrumente in der neuropsychologischen Leistungsdiagnostik. Er erfasst neben motorischen und visuellen Basisfähigkeiten insbesondere Facetten der Aufmerksamkeits- und Exekutivfunktionen. Der TMT hat sich als besonders sensitiv bei der Detektion von Leistungseinbußen bei Patienten mit leichter kognitiver Störung sowie bei Demenz mit Alzheimerkrankheit erwiesen und wird daher regelmäßig im diagnostischen Prozess eingesetzt.

Eine wichtige Limitation beim Einsatz der Originalversion des TMT besteht darin, dass die Anwendung des Untertests B eine genaue Kenntnis der alphabetischen Buchstabenfolge voraussetzt. Um diesem Problem zu begegnen, wurden Varianten des TMT entwickelt, die unter Zuhilfenahme verschiedener Methoden auf Schriftzeichen als Stimuli verzichten. Unter den in Abschnitt 1.8. vorgestellten Verfahren werden von den Autoren selbst der *Color Trails Tests 2*, der *Shape Trail Test B* und der *Trail Making Test Black & White B* als Alternativen zum TMT-B bezeichnet. Diese Äquivalenz ist konzeptuell fraglich, da alle genannten Verfahren im Unterschied zum TMT-B zusätzliche Distraktoren einführen. Darüber hinaus wird die Äquivalenz der Verfahren nur teilweise durch die Forschungsbefunde gedeckt: Beim CTT-2

sprechen die teils niedrigen Korrelationen und die fehlende Diskriminierbarkeit zwischen gesunden Personen und Patienten mit MCI gegen eine Eignung als Alternative zum TMT-B. Beim STT-B fehlen Angaben zur Konstruktvalidität im Sinne einer Korrelation zum TMT-B. Die ersten Befunde zum TMT-B&W-B scheinen vielversprechend, allerdings wurde das Verfahren bisher erst in einer Untersuchung an wenigen Personen eingesetzt.

In der vorliegenden Arbeit wird mit dem Berliner Shape Trail Test (B-STT) eine weitere Variante des Trail Making Tests untersucht. Der B-STT wurde im Jahr 2009 vom Verfasser dieser Arbeit unter dem Arbeitstitel Shape Trail Test entwickelt; die spätere Ergänzung des Namens erfolgte aufgrund der Veröffentlichung des chinesischen Shape Trail Tests im Jahr 2013. Ziel des B-STT ist es, exekutive Funktionen in ähnlicher Weise wie der TMT-B zu erfassen, dabei jedoch auf die Verwendung von Schriftzeichen zu verzichten. Zusätzlich sollen die räumlichen Eigenschaften der Item-Platzierung und damit die Anforderungen an motorische und visuelle Fähigkeiten an die des TMT-A angeglichen werden. Dies soll eine stärkere Isolierbarkeit der exekutiven Komponente ermöglichen und gleichzeitig die Ökonomie des Verfahrens durch eine schnellere Durchführbarkeit erhöhen. Ein relevanter Unterschied im Vergleich zum CTT-2, STT-B und TMT-B&W-B besteht zudem darin, dass bewusst auf Distraktoren verzichtet wird, so dass die Itemzahl der des TMT-B entspricht.

Die allgemeine Fragestellung lautet, ob der Berliner Shape Trail Test unter Berücksichtigung psychodiagnostischer Testgütekriterien ein geeignetes Verfahren zur neuropsychologischen Leistungsdiagnostik und darüber hinaus eine angemessene Alternative zum Trail Making Test B darstellt, die bei vergleichbaren diagnostischen Klassifikationsleistungen unabhängig von Kenntnissen des lateinischen Alphabets anwendbar ist. Weiterhin sollen zur Vorbereitung einer späteren Normierung die Einflüsse demografischer Faktoren untersucht werden.

Zur Beantwortung dieser Fragen wurde der B-STT in zwei Untersuchungen eingesetzt, die die ersten Anwendungen des Verfahrens darstellten. In beiden Teilstudien war die Bearbeitungszeit der interessierende Leistungsindex. In einer Untersuchung an gesunden Probanden wurden die Einflüsse demografischer Faktoren herausgestellt, und in einer Teilstichprobe wurde die Zuverlässigkeit der Messung über die Zeit (Test-Retest-Reliabilität) untersucht. Im zweiten Teil, einer klinischen Validierungsstudie, wurde untersucht, inwieweit das Verfahren zwischen gesunden Probanden, Patienten mit leichter kognitiver Störung und Patienten mit Demenz bei Alzheimerkrankheit (jeweils ohne neurologische oder psychiatrische Komorbidität) zu diskriminieren vermag (Kriteriumsvalidität). Teil beider Untersuchungen war zusätzlich eine

Analyse des Zusammenhangs mit dem Trail Making Test, wodurch die Konstruktvalidität des Verfahrens geprüft wurde. Am gesamten Datensatz sollte zudem untersucht werden, ob beim B-STT insgesamt eine kürzere Bearbeitungszeit benötigt wird als beim TMT-B.

Für die allgemeinen Fragestellungen wurden unter Berücksichtigung der Literaturbefunde und aufgrund theoretischer Überlegungen folgende konkrete Forschungshypothesen formuliert:

#### *Einfluss demografischer Parameter*

Hypothese 1: Die Bearbeitungszeit im B-STT ist alters- und bildungsabhängig. Ein höheres Lebensalter und ein geringeres Bildungsniveau stehen dabei mit einer längeren Bearbeitungszeit beim B-STT in Verbindung.

Bezüglich des Einflusses des Geschlechts auf die Bearbeitungszeit im B-STT wurde keine gerichtete Hypothese formuliert, da die Literatur zum TMT darauf verwies, dass keine relevanten Geschlechtsunterschiede vorliegen.

#### *Test-Retest-Reliabilität*

Hypothese 2: Es besteht ein positiver Zusammenhang zwischen den Bearbeitungszeiten des B-STT von zwei zeitlich aufeinanderfolgenden Messungen.

#### *Validität*

Hypothese 3: Es wird angenommen, dass (a) die Bearbeitungszeit im B-STT mit den Bearbeitungszeiten im TMT-A und dem TMT-B korreliert und dass (b) die Korrelation zwischen dem B-STT und dem TMT-B größer ist als die zwischen B-STT und TMT-A.

Hypothese 4: Patienten mit leichter kognitiver Störung (MCI) und Patienten mit Demenz bei Alzheimerkrankheit (DAT) unterscheiden sich in ihren Leistungen im B-STT voneinander und von gesunden Probanden der gleichen Altersgruppe (KG), wobei KG die höchsten und DAT die geringsten Leistungen zeigen.

#### *Testökonomie*

Hypothese 5: Die Bearbeitungszeiten von B-STT und TMT-B unterscheiden sich. Die Bearbeitung des B-STT erfolgt dabei schneller als die des TMT-B.

---

## 2. Methodik

### 2.1. *Untersuchungskollektiv*

#### 2.1.1. *Stichprobe der gesunden Probanden*

Für die demografische Charakterisierung des Berliner Shape Trail Test (B-STT) wurden 235 gesunde Probanden im Alter von 21–80 Jahren rekrutiert. Als Ausschlusskriterien galten (1) das Vorliegen neurologischer oder psychiatrischer Erkrankungen, (2) Schädel-Hirn-Traumata in der Anamnese, (3) die Einnahme psychotroper Substanzen, (4) schwere Einschränkungen der Handmotorik oder Sehfähigkeit und (5) Durchführung des Trail Making Tests vor weniger als zwölf Monaten. Die Rekrutierung erfolgte dabei auf drei Wegen:

Eingeschlossen wurden erstens 113 Personen, die sich aufgrund subjektiver kognitiver Einbußen zur Diagnostik in der Gedächtnissprechstunde der Psychiatrischen Universitätsklinik der Charité im St. Hedwig-Krankenhaus vorgestellt hatten, bei denen jedoch psychiatrische und neurologische Erkrankungen ausgeschlossen werden konnten. Die Einschätzung des Zustandes erfolgte im Rahmen einer interdisziplinären Konsensuskonferenz unter Beteiligung erfahrener Psychiater und Neuropsychologen. Berücksichtigt wurden dabei strukturiert erhobene eigen- und fremdanamnestische Angaben, der psychopathologische Befund, eine ausführliche neuropsychologische Leistungsuntersuchung (siehe Abschnitt 2.2.2.) sowie paraklinische Befunde einschließlich kranialer Bildgebung und Labordiagnostik. Bei diesem Teil der Stichprobe wurde der B-STT mit Einwilligung der Versuchspersonen als Teil des neuropsychologischen Assessments durchgeführt.

Zweitens erfolgte eine Rekrutierung von 60 Personen durch Flyer, Aushänge und Anzeigen sowie durch mündliche Ansprache von Angehörigen von Patienten. Die Prüfung der Ausschlusskriterien erfolgte durch strukturierte Erhebung der Anamnese mit Erfragung der demografischen Daten, des Bildungsweges, der psychiatrischen und somatischen Krankenvorgeschichte, der Medikation sowie des Substanzgebrauchs. Bei allen Probanden, die zum Zeitpunkt der Untersuchung älter als 50 Jahre waren, wurde der Mini-Mental-Status-Test (MMST; [99]) durchgeführt; ein Gesamtwert von mindestens 27 Punkten war Voraussetzung für den Einschluss in die Studie. Jeweils 30 Probanden wurden dabei durch zwei Psychologiestudentinnen im Rahmen von Studienprojekten erhoben. Die

Durchführungsobjektivität wurde zuvor durch ausführliches Training des Untersuchungsablaufs sichergestellt.

Der dritte Teil der Stichprobe, bestehend aus 62 Personen, wurde im Rahmen einer Studie zum Arbeitsgedächtnistraining [100] erhoben. Für die aktuelle Untersuchung fanden dabei ausschließlich Daten Verwendung, die vor Trainingsbeginn gewonnen wurden. Das Assessment beinhaltete neben einer strukturierten Erhebung der Anamnese auch eine ausführliche neuropsychologische Untersuchung mit der Batterie CERAD-Plus [11]. Bei 19 Personen, die im Rahmen der Trainingsstudie der Kontrollgruppe ohne Training zugeordnet waren, wurden zusätzlich die Daten der Verlaufsmessung nach vier Wochen für die Bestimmung der Test-Retest-Reliabilität des B-SST verwendet. Die Anzahl der Personen ist unter Annahme einer mindestens starken Korrelation ( $r \geq .60$ ), eines Fehlers erster Art von 5 % und eines Fehlers zweiter Art von 20 % ausreichend [101].

### *2.1.2. Stichprobe in der klinischen Studie*

Für die klinische Erprobung des B-STT wurden 36 Patienten mit Demenz bei Alzheimerkrankheit (DAT) und 82 Patienten mit leichter kognitiver Störung (MCI), jeweils ohne komorbide neurologische oder psychiatrische Erkrankungen, rekrutiert. Eine Subtypenklassifizierung der MCI erfolgte im Rahmen dieser Untersuchung nicht. Die Patientengruppen wurden in der Gedächtnissprechstunde sowie in der Gerontopsychiatrischen Tagesklinik der Psychiatrischen Universitätsklinik der Charité im St. Hedwig-Krankenhaus rekrutiert. Der B-STT wurde hierbei mit Einwilligung der Patienten als Teil der neuropsychologischen Leistungsdiagnostik durchgeführt. Kriterien für den Einschluss in die Studie waren (1) ein Alter von 65–80 Jahren, (2) die Diagnose einer klinisch wahrscheinlichen beginnenden Demenz bei Alzheimerkrankheit (DAT) oder einer leichten kognitiven Störung (MCI) ohne neurologische oder psychiatrische Komorbidität und (3) das Absolvieren der neuropsychologischen Testbatterie einschließlich des TMT und des B-STT.

Die Diagnosestellung erfolgte im Rahmen einer interdisziplinären Konsensuskonferenz nach dem unter 2.1.1. beschriebenen Verfahren. Die Ergebnisse des B-STT wurden nicht für die Diagnosestellung berücksichtigt. Da die Diagnosestellung erst nach Durchführung der neuropsychologischen Untersuchung erfolgen konnte, wurden auch Patienten mit anderen Störungsbildern mit dem B-STT untersucht. Diese Population ist nicht Gegenstand der vorliegenden Arbeit.

Den Gruppen der Patienten mit DAT und MCI wurde eine Kontrollgruppe (KG) von 107 gesunden Personen gegenübergestellt. Dabei handelt es sich um alle aus der Stichprobe der gesunden Probanden stammenden Personen, deren Alter 65–80 Jahre betrug.

## 2.2. Untersuchungsinstrumente

Für die Untersuchung wurde der B-STT in eine Batterie weiterer Testverfahren eingebettet. Dabei handelt es sich um Untertests der erweiterten neuropsychologischen Batterie des *Consortium to Establish a Registry for Alzheimer's Disease* (CERAD-Plus; [11]). Eine Untersuchung mit allen Instrumenten erfolgte dabei regelhaft bei den Patienten; lediglich bei zwei Patienten wurde aus Zeitgründen der Test *Wortflüssigkeit S-Wörter* nicht durchgeführt. In der Gruppe der gesunden Probanden wurden, sofern nicht in der Gedächtnissprechstunde rekrutiert, aus Gründen der Ökonomie teilweise nicht alle Untersuchungen durchgeführt. Eine Übersicht der verwendeten Testverfahren in Abhängigkeit von der Stichprobenzugehörigkeit ist in Tabelle 1 dargestellt. Angegeben ist die Anzahl der Personen, bei der das jeweilige Verfahren durchgeführt wurde.

**Tabelle 1: In der Studie durchgeführte neuropsychologische Testverfahren**

Testverfahren	Patientengruppen		gesunde Probanden*	
	MCI, DAT		21–50 Jahre	51–80 Jahre
	Anzahl = 118		Anzahl = 59	Anzahl = 176
Wortflüssigkeit Tiere	118	0	145	
modif. Boston Naming Test	118	0	145	
Mini-Mental-Status-Test	118	0	176	
Wortliste Lernen	118	0	145	
Figuren Abzeichnen	118	0	145	
Wortliste Abrufen	118	0	145	
Wortliste Wiedererkennen	118	0	145	
Figuren Abrufen	118	0	145	
Trail Making Test A	118	59	176	
Trail Making Test B	118	59	176	
Berliner Shape Trail Test	118	59	176	
Wortflüssigkeit S-Wörter	116	0	145	

\* Die getrennte Darstellung der Altersbereiche verdeutlicht, dass Probanden im Altersbereich von 21–50 Jahren ausschließlich mit dem Trail Making Test und dem Berliner Shape Trail Test untersucht wurden.

Die Reihenfolge, in der die Testverfahren durchgeführt wurden, entspricht der Sortierung innerhalb der Tabelle. So wurde der B-STT in jedem Fall im Anschluss an TMT-A und TMT-B absolviert.

### 2.2.1. Berliner Shape Trail Test (B-STT)

#### *Testentwicklung und -eigenschaften*

Die Konzeption und Konstruktion des B-STT erfolgte durch den Verfasser der vorliegenden Arbeit in Anlehnung an den Color Trails Tests (CTT; [81]). Wie beim CTT-2 besteht das Stimulusmaterial beim B-STT lediglich aus Zahlen, die auf der Testvorlage jeweils zweifach vorkommen. Der B-STT unterscheidet sich jedoch in drei relevanten Aspekten vom CTT-2: Erstens wird auf die Verwendung von Farben verzichtet; die Zahlen befinden sich stattdessen in Kreisen und Quadraten. Dies ermöglicht, anders als beim CTT, den kostengünstigen Ausdruck der Testform. Zweitens wird auf den Einsatz von zusätzlichen Distraktorreizen verzichtet, wodurch die Zahl der Stimuli dem TMT angeglichen wird. Bedingt durch die Variation der Aufgabenstellung (s. u.) ergibt sich ein dritter Unterschied: Das höchste Zeichen auf der Testvorlage ist die Zahl 13.

Wie beim TMT befindet sich das Stimulusmaterial in einem scheinbar willkürlichem Muster auf einem Arbeitsblatt vom Format DIN A4 (210 mm × 297 mm; Anhang 5.1.). Aufgrund des Befundes, dass das Messergebnis beim TMT-B durch die längere Gesamtstrecke und die größere räumliche Komplexität überlagert ist [29], erfolgte bei der Positionierung der Zielreize im B-STT eine stärkere Orientierung am TMT-A, um die Messung um diese Störvariablen zu bereinigen. Konkret wurden die folgenden Zielkriterien verwendet:

*Längenkriterium:* Die Gesamtlänge beim Verbinden der Stimuli sowie die minimale und die maximale Distanz zwischen zwei aufeinanderfolgenden Stimuli sollten mit den entsprechenden Werten im TMT-A vergleichbar sein.

*Distraktorkriterium:* Die Zahl der Distraktoren in der Umgebung des aktuellen Zeichens sollte ähnlich groß sein wie im TMT-A. Als Distraktor wurde hierbei für die Testkonstruktion jedes Zeichen gewertet, welches (1) nicht den nächsten Zielreiz darstellt und (2) dessen Distanz zum aktuellen Zeichen mindestens 0,5 cm geringer ist als die Distanz zwischen dem aktuellen Zeichen und dem nachfolgenden Zielreiz. So liegen zum Beispiel beim TMT-A die Zahlen 3, 6 und 24 näher am Startzeichen 1 als die eigentlich nachfolgende Zahl 2. Bezüglich der

Distraktoren wurde weiterhin unterschieden zwischen der Gesamtzahl der nach den genannten Kriterien gezählten Zeichen (Distraktor Typ A) und der Anzahl der Zeichen, die nicht bereits mit einem anderen Zeichen verbunden sind bzw. für deren Erreichen keine Verbindungslinie zwischen zwei Zeichen überkreuzt werden muss (Distraktor Typ B).

Die Eigenschaften des Stimulusmaterials von TMT-A, TMT-B und B-STT sind in Tabelle 2 wiedergegeben. Ersichtlich wird, dass den geforderten Kriterien bei der Konstruktion des B-STT weitgehend entsprochen werden konnte.

**Tabelle 2: Räumlich-strukturelle Eigenschaften von TMT-A, TMT-B und B-STT**

Eigenschaft	Testverfahren		
	TMT-A	TMT-B	B-STT
Anzahl der Zeichen	25	25	25
Gesamtlänge*	139,0	181,9	140,6
Distanz Min/Max*	2,5/11,1	3,1/15	2,8/10,3
Distanz M (SD)*	5,8 (2,7)	7,6 (3,3)	5,9 (2,0)
Distraktor Typ A Summe	101	154	102
Distraktor Typ A Min/Max	0/13	1/19	0/11
Distraktor Typ A M (SD)	4,2 (3,8)	6,4 (4,6)	4,3 (3,2)
Distraktor Typ B Summe	34	52	34
Distraktor Typ B Min/Max	0/6	0/12	0/9
Distraktor Typ B M (SD)	1,4 (1,7)	2,2 (2,9)	1,4 (2,2)

\* Angaben in cm; Längenmessungen erfolgten jeweils vom Zentrum der Kreise bzw. Quadrate aus

### *Durchführung des B-STT*

Die Aufgabe der Testperson besteht darin, die Zahlen aufsteigend so schnell wie möglich so miteinander zu verbinden, dass jeweils zweimal nacheinander die gleiche Zahl kommt und zwischen Kreisen und Quadraten alterniert wird (Kreis-1–Quadrat-1–Kreis-2–Quadrat-2–...). Die wörtliche Testinstruktion befindet sich in Anhang 5.2.; zur besseren Verständlichkeit wird in der Anweisung der Begriff *Viereck* anstelle von *Quadrat* verwendet. Den Probanden wird zunächst eine Übungsversion mit acht Zahlen vorgelegt (Abbildung 1), anhand derer die Aufgabe erläutert wird. Auftretende Fehler während der Übung werden korrigiert, die Instruktion gegebenenfalls wiederholt. Wurde die Übung verstanden und erfolgreich absolviert, erfolgt die Durchführung der eigentlichen Untersuchung. Die Zeitmessung beginnt mit dem Vorlegen des Arbeitsblattes und Deuten des Untersuchers auf das erste Zeichen (Kreis-1) und endet mit dem Erreichen des letzten Zeichens (Kreis-13). Beim Auftreten von Fehlern wird unmittelbar darauf hingewiesen,

ohne die Lösung zu offenbaren. Leistungsparameter für die vorliegende Untersuchung ist die Bearbeitungszeit. Eine maximale Bearbeitungsdauer wurde nicht festgelegt. Eine Dokumentation der Fehlerzahl und Fehlertypen erfolgt zwar, die Auswertung wird jedoch in einer separaten Arbeit durchgeführt.

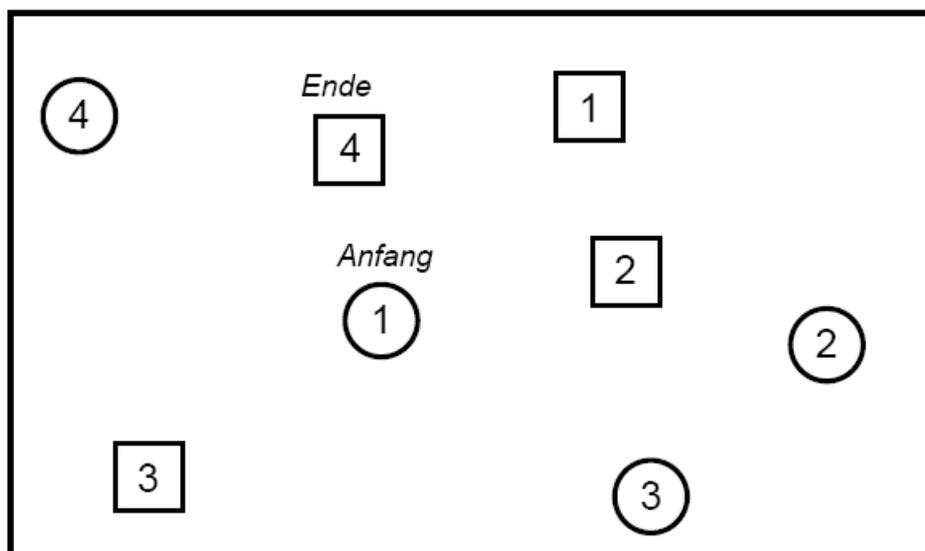


Abbildung 1: Übungsversion des B-STT (Originalgröße)

### 2.2.2. Neuropsychologische Batterie CERAD-Plus

Die neuropsychologische Batterie des *Consortium to Establish a Registry for Alzheimer's Disease* (CERAD-NP; [102]) ist eine Sammlung neuropsychologischer Untersuchungsverfahren zur Demenzdiagnostik, die eine besondere Eignung bei der Feststellung typischer Leistungsminderungen bei der DAT aufweisen. Im Zuge der deutschsprachigen Übersetzung durch die Arbeitsgruppe der Memory Clinic Basel [11] wurde die Batterie zur Verbesserung der Messeigenschaften bei subkorticalen Störungen um weitere Testverfahren (Trail Making Test A und B, Wortflüssigkeit S-Wörter) zur CERAD-Plus erweitert. Die Untertests der CERAD-Plus werden in den folgenden Abschnitten vorgestellt.

#### *Wortflüssigkeit Tiere und S-Wörter*

Bei den beiden Untertests zur Wortflüssigkeit besteht die Aufgabe der Probanden darin, anhand einer vorgegebenen semantischen bzw. lexikalischen Kategorie (Tierarten bzw. Wörter mit S) innerhalb einer Minute möglichst viele verschiedene Wörter zu produzieren. Gewertet wird die Anzahl der korrekten Antworten ohne Wiederholungen. Mit diesen Aufgaben werden

---

sprachliche und exekutive Funktionen geprüft. Die Wortflüssigkeit für Tierarten ist zudem mit dem semantischen Altgedächtnis assoziiert [11].

### *Modifizierter Boston Naming Test*

Beim modifizierten Boston Naming Test werden dem Probanden fünfzehn Strichzeichnungen vorgelegt, auf denen verschiedene Objekte abgebildet sind. Diese sollen korrekt benannt werden, wobei die gesuchten Wörter nach der Häufigkeit ihres Auftretens in der amerikanischen Sprache in drei Kategorien (häufig, mittel, selten) unterteilt werden können. Es können maximal fünfzehn Punkte erreicht werden. Für die Bewertung steht ein Lösungsschlüssel zur Verfügung. Die Aufgabe quantifiziert die Fähigkeit zur Objekterkennung und -benennung [11].

### *Mini-Mental-Status-Test*

Der Mini-Mental-Status-Test ist ein Screening-Instrument zur Detektion kognitiver Defizite, kann aber auch zur Verlaufsbeurteilung herangezogen werden. Er beinhaltet insgesamt 30 Items, die die zeitliche und örtliche Orientierung, die Merk- und Konzentrationsfähigkeit, die konstruktive Praxis sowie sprachliche Fähigkeiten erfassen [11,99].

### *Wortliste Lernen, Abrufen und Wiedererkennen*

Bei den Wortlistenaufgaben werden dem Probanden zunächst zehn unrelatierte Wörter in drei Lerndurchgängen mit unterschiedlicher Sortierung vorgelegt, wobei die Wörter vom Probanden laut vorgelesen und dabei gelernt werden sollen. Dokumentiert werden dabei die Abrufleistung bei jedem Lerndurchgang sowie das Gesamtlernvolumen, welches sich aus der Summe der einzelnen Abrufleistungen ergibt. Nach Absolvierung der nachfolgenden Aufgabe *Figuren Abzeichnen* werden ein weiterer Abrufdurchgang sowie eine Rekognitionsaufgabe durchgeführt, bei der die zehn Zielwörter von zehn Distraktoren unterschieden werden sollen. Ein Punkt wird jeweils für jedes richtig genannte bzw. korrekt als bekannt oder neu klassifizierte Wort erteilt. Damit können bei den Abrufdurchgängen bis zu zehn Punkte und bei der Rekognitionsaufgabe bis zu 20 Punkte erzielt werden. Mit den Wortlistenaufgaben werden verschiedene Aspekte der verbalen Neugedächtnisbildung, insbesondere Speicher- und Zugriffsprozesse, untersucht [11].

### *Figuren Abzeichnen und Abrufen*

Die Anforderung der Aufgabe Figuren Abzeichnen besteht darin, vier geometrische Figuren ansteigender Komplexität möglichst sorgfältig abzuzeichnen. Der verzögerte Abruf der Figuren wird nach der Aufgabe *Wortliste Wiedererkennen* durchgeführt. Hierbei wird der Proband aufgefordert, die zuvor gezeichneten Figuren aus dem Gedächtnis möglichst sorgfältig wiederzugeben. Die Qualität der Ausführung wird in beiden Konditionen anhand von elf Kriterien durch den Untersucher bewertet. Mit diesen Aufgaben werden die visuokonstruktiven Fähigkeiten sowie die Neugedächtnisbildung für nonverbales Material untersucht [11].

### *Trail Making Test A und B*

Die Durchführung des TMT und die erfassten kognitiven Funktionen wurden bereits ausführlich in den Abschnitten 1.2. und 1.3. erläutert. Von der Standarddurchführung wurde im Rahmen dieser Untersuchung insofern abgewichen, als hinsichtlich der Bearbeitungszeit kein Abbruchkriterium vorlag. Dadurch konnten alle Probanden die Aufgabe bis zum letzten Zeichen absolvieren. Ein ähnliches Vorgehen wurde beim B-STT gewählt. Dies sollte gewährleisten, dass hinsichtlich des Leistungsindikators Bearbeitungszeit das gesamte Fähigkeitspektrum der Stichprobe abgebildet werden kann. Die in der Studie verwendete Testinstruktion befindet sich in Anhang 5.2.

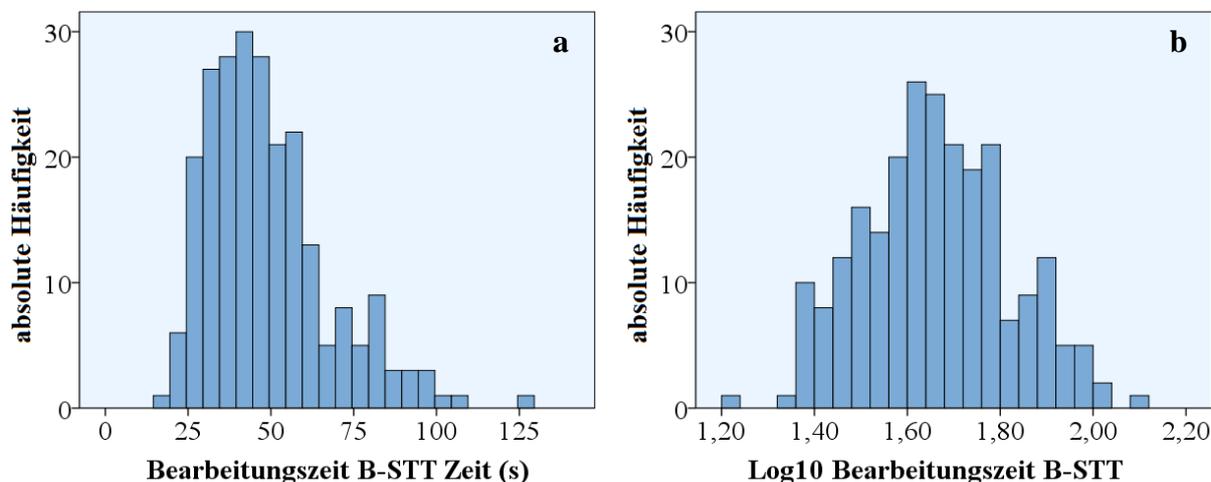
## **2.3. Statistische Methoden**

Die statistische Auswertung wurde mit dem Programm IBM SPSS Statistics 21 durchgeführt. In einigen Fällen wurden zusätzliche Programme verwendet. Diese sind bei den entsprechenden Analysen in den folgenden Abschnitten benannt.

### *2.3.1. Auswahl der statistischen Verfahren*

Die Analysen wurden, sofern die statistischen Voraussetzungen erfüllt waren, mit parametrischen Verfahren durchgeführt. Die Prüfung der Voraussetzungen mit Angabe der Testgrößen ist im Ergebnisteil bei den einzelnen Analysen dokumentiert. Bei Verletzung der Voraussetzungen für parametrische Verfahren wurde auf nichtparametrische Verfahren zurückgegriffen. Ausgenommen davon sind Fälle einer solitären Verletzung der bei parametrischen Verfahren vorausgesetzten Normalverteilung der Daten, Differenzwerte oder Residuen. Hier wurde eine Normalisierung mithilfe einer  $\text{Log}_{10}$ -Transformation durchgeführt.

Für die Bearbeitungszeit im B-STT in der Gesamtstichprobe der gesunden Probanden ist die  $\text{Log}_{10}$ -Transformation in Abbildung 2 demonstriert. Erkennbar sind die deutliche Rechtsschiefe vor der Transformation und die Annäherung an die Normalverteilung nach der Transformation.



**Abbildung 2:** Häufigkeitsverteilung der Bearbeitungszeit im B-STT (a) vor und (b) nach  $\text{Log}_{10}$ -Transformation

### 2.3.2. Festlegung des Signifikanzniveaus und Korrektur des Alpha-Fehlers

Das Signifikanzniveau wurde für alle im Folgenden genannten Analysen auf  $p=.050$  festgelegt. Alle Analysen erfolgten zweiseitig. Aufgrund der Vielzahl der untersuchten Fragestellungen und verwendeten Testverfahren besteht die Gefahr einer Zunahme des Fehlers erster Art. Um diesem zu begegnen, wurden die p-Werte der Testverfahren, die in der Inferenzstatistik Anwendung fanden, mithilfe der Bonferroni-Holm-Methode adjustiert. Da zwischen der Studie zur demografischen Charakterisierung und der klinischen Studie eine Schnittmenge von 107 Personen vorlag, wurde die Adjustierung konservativ über die Gesamtpopulation beider Gruppen durchgeführt.

Die Berechnung der angepassten p-Werte erfolgte manuell mithilfe einer Excel-Matrix. Berücksichtigt wurden alle p-Werte der inferenzstatistischen Analysen mit Ausnahme der nach signifikanten Globaltests (Rangvarianzanalyse) durchgeführten Paarvergleiche. In solchen Fällen erfolgte die Adjustierung direkt auf Ebene der Post-hoc-Prüfungen.

Analysen, die nicht zur Hypothesenprüfung dienten, wurden nicht bei der globalen Adjustierung berücksichtigt. Dies betrifft Analysen, die zur Beschreibung der Stichprobeneigenschaften erfolgten, sowie die deskriptive Korrelationsmatrix der neuropsychologischen Verfahren in der klinischen Studie.

### 2.3.3. Charakterisierung der Stichproben

#### *Demografische Eigenschaften der Untersuchungsgruppen*

Die statistischen Kennwerte Mittelwert und Standardabweichung für Alter und Bildungsniveau wurden bei den gesunden Probanden differenziert für sechs Altersgruppen (21–30, 31–40, 41–50, 51–60, 61–70, 71–80 Jahre) bestimmt. Zusätzlich wurde die Geschlechterverteilung ausgezählt. In ähnlicher Weise wurden die Eigenschaften der Gruppen in der klinischen Studie getrennt nach der Gruppenzugehörigkeit (gesunde Kontrollgruppe [KG], Patienten mit leichter kognitiver Störung [MCI], Patienten mit Demenz bei Alzheimerkrankheit [DAT]) berechnet.

Im Sinne einer beschreibenden Charakterisierung der Stichprobe erfolgte in beiden Teilstudien weiterhin eine Unterschiedsprüfung der Gruppen bezüglich der demografischen Parameter. Die Untersuchung erfolgte für das Alter mithilfe einer Varianzanalyse mit anschließenden Post-hoc-t-Tests für unabhängige Stichproben. Für das Bildungsniveau wurde der Kruskal-Wallis-Test, ein nichtparametrisches Analogon zur Varianzanalyse, eingesetzt. Der Vergleich der Geschlechterverteilungen wurde mit einem Chi-Quadrat-Test durchgeführt.

#### *Testleistungen bei den gesunden Probanden*

Für die Bearbeitungszeiten im B-STT sowie im TMT-A und TMT-B wurden die statistischen Kennwerte Mittelwert, Median, Standardabweichung, Minimum und Maximum differenziert für sechs Altersgruppen (21–30, 31–40, 41–50, 51–60, 61–70, 71–80 Jahre) errechnet. Zusätzlich wurden die Bearbeitungszeiten des B-STT für die 19 Personen, die die Tests zweimalig absolvierten, getrennt für den ersten und zweiten Messzeitpunkt bestimmt.

#### *Testleistungen in der klinischen Studie*

Für die Bearbeitungszeit im B-STT wurden die statistischen Kennwerte Mittelwert, Median, Standardabweichung, Minimum und Maximum getrennt für die Gruppen KG, MCI und DAT berechnet. Die Gruppen wurden außerdem bezüglich aller weiteren durchgeführten neuropsychologischen Testverfahren der CERAD-Plus charakterisiert, wobei eine Bestimmung von Mittelwert und Standardabweichung erfolgte. Die Gruppen wurden anschließend auf lokalem Signifikanzniveau bezüglich der Unterschiede in den einzelnen Testverfahren der CERAD-Plus untersucht. Bei dieser Untersuchung wurden alle Personen eingeschlossen, für die ein Testrohwert für das jeweilige Verfahren vorlag (paarweiser Fallausschluss). Abhängig vom

Zutreffen der Voraussetzungen wurden die ANOVA oder der Kruskal-Wallis-Test durchgeführt; die anschließenden Paarvergleiche erfolgten mit dem t-test für unabhängige Stichproben bzw. dem Wilcoxon-Mann-Whitney-Test.

#### *2.3.4. Einfluss demografischer Faktoren auf die Bearbeitungszeit im B-STT*

Der Zusammenhang zwischen dem Alter bzw. der Bildung und der Bearbeitungszeit im B-STT wurde zunächst korrelativ bestimmt (Pearson-Korrelationskoeffizient  $r$ ). Die eigentlich notwendige Voraussetzung der Normalverteilung wurde in diesem Fall nicht geprüft, da aufgrund der Stichprobengröße eine ausreichende Robustheit vorausgesetzt werden konnte [103]. Für den Einfluss von Alter und Bildung auf die Bearbeitungszeit wurde außerdem nach Prüfung der Voraussetzungen eine Regressionsanalyse mit Alter und Bildungsniveau als unabhängige Variablen und dem dekadischen Logarithmus der Bearbeitungszeit als abhängige Variable erstellt. Der Einfluss des Geschlechts wurde mittels Kovarianzanalyse untersucht; hierbei wurde für Alter und Bildung kontrolliert.

#### *2.3.5. Test-Retest-Reliabilität des B-STT*

Die Test-Retest-Reliabilität der Bearbeitungszeiten des B-STT wurde nach Prüfung der Normalverteilung als Pearson-Korrelation zwischen den Testwerten der beiden Messzeitpunkte in der Reliabilitätsstudie berechnet.

#### *2.3.6. Validität*

##### *Konstruktvalidität*

Die Untersuchung der Konstruktvalidität des B-STT erfolgte in zwei Schritten: Zunächst wurden die Pearson-Korrelationskoeffizienten für Bearbeitungszeiten im B-STT und TMT-A sowie im B-STT und TMT-B bestimmt; hierbei wurde erneut davon ausgegangen, dass aufgrund der Stichprobengröße eine ausreichende Stabilität des Korrelationskoeffizienten vorliegt (Vergleich Abschnitt 2.3.4.). Anschließend wurden die Korrelationskoeffizienten einer Unterschiedsprüfung unterzogen. Hierfür wurde CorrTest2 verwendet, ein online verfügbares Programm zum Vergleich von Korrelationen bei abhängigen Messungen unter Nutzung der Fisher-z-Transformation [104]. Die Bestimmung wurde getrennt für die Studie zur demografischen Charakterisierung und die klinische Studie durchgeführt, um festzustellen, ob die ermittelten

Zusammenhänge sowohl im Wertebereich kognitiv unbeeinträchtigter Personen als auch über den gesamten Fähigkeitsbereich der Bearbeitungszeit vorliegen.

In der klinischen Stichprobe wurde weiterhin geprüft, ob die beobachtete Assoziation zwischen den Leistungen im B-STT und TMT-B bzw. TMT-A nicht nur auf Zusammenhänge zwischen den spezifischen Testverfahren, sondern auf eine Abnahme des globalen kognitiven Funktionsniveaus zurückzuführen sind. Der Einfluss der allgemeinen kognitiven Leistungsfähigkeit, operationalisiert durch den Punktwert im Mini-Mental-Status-Test (MMST), wurde durch partielle Korrelation eliminiert. Die ermittelten Korrelationskoeffizienten wurden danach erneut mittels CorrTest2 miteinander verglichen.

Der letzte Teil der Untersuchungen zur Konstruktvalidität bestand in einer deskriptiven Bestimmung der Stärke des Zusammenhangs der Leistungsindikatoren des B-STT zu den weiteren neuropsychologischen Verfahren. Nach dem Verfahren des paarweisen Fallausschlusses wurden jeweils alle Personen im Alter von 65–80 Jahren eingeschlossen, bei denen die jeweiligen Testverfahren durchgeführt worden waren (S-Wörter 213 Personen, alle anderen Verfahren 215 Personen). Zum Ausschluss einer Überlagerung des Zusammenhangs durch das globale kognitive Funktionsniveau wurden die Zusammenhänge als partielle Korrelationen kontrolliert für den Summenwert des MMST berechnet.

### *Kriteriumsvalidität*

Die Analysen bezüglich der Gruppenunterschiede in der Bearbeitungszeit im B-STT wurden zunächst mithilfe des Kruskal-Wallis-Tests durchgeführt. Darüber hinaus wurden als Maß für die Diskriminierbarkeit der Gruppen die Effektgrößen (Cohens  $d$ ) der Gruppenunterschiede bestimmt. Da die Daten nicht normalverteilt waren, wurde ein  $d$ -Schätzer für nonparametrische Daten verwendet, der gegenüber Abweichungen von der Normalverteilung robuster ist [105]. Zur Ermittlung der Trennschärfe der Bearbeitungszeit im B-STT wurden *Receiver-Operating-Characteristic*-Analysen (ROC-Analysen) für den B-STT durchgeführt und denen des TMT-B gegenübergestellt. Ermittelt wurde jeweils die Fläche unter der ROC-Kurve (*area under the curve*; AUC), die bei der Gegenüberstellung von Sensitivität und Spezifität der Leistungsparameter für jeden Punktwert im beobachteten Merkmalskontinuum unter Berücksichtigung der vordefinierten Gruppen berechnet wurde.

### *2.3.7. Unterschiede in den Bearbeitungszeiten von B-STT und TMT-B*

Die Bearbeitungszeiten von B-STT und TMT-B wurden mit dem Wilcoxon-Vorzeichen-Rang-Test verglichen. Anschließend wurde mit dem Kruskal-Wallis-Test und anschließenden Post-hoc-Vergleichen mit Wilcoxon-Mann-Whitney-Tests geprüft, ob die Zeitunterschiede zwischen den Verfahren in den verschiedenen Untersuchungspopulationen (Kontrollgruppe < 65 Jahre, Kontrollgruppe  $\geq$  65 Jahre, Patienten mit MCI, Patienten mit DAT) unterschiedlich ausgeprägt sind.

### 3. Ergebnisse

#### 3.1. Deskriptive Darstellung der Untersuchungsergebnisse

##### 3.1.1. Demografische Eigenschaften der Untersuchungsgruppen

###### Gesunde Probanden

Ein Überblick über die Altersverteilung sowie demografische Charakteristika der Gruppe der gesunden Probanden wird in Tabelle 3 gewährt. Neben Angaben zur Anzahl der Teilnehmer (n) und Geschlechtsverteilung finden sich jeweils die Mittelwerte (M) und Standardabweichungen (SD) von Alter und Bildung. Die angegebene Bildungsdauer berechnet sich als Summe der Dauer der Schulzeit ohne Klassenwiederholungen und der Dauer der Ausbildung bzw. des Studiums mit dem höchsten Abschluss.

**Tabelle 3: Demografische Eigenschaften der gesunden Probanden getrennt nach sechs Altersgruppen**

	Altersgruppe					
	21–30 (n = 36)	31–40 (n = 12)	41–50 (n = 11)	51–60 (n = 34)	61–70 (n = 82)	71–80 (n = 60)
Geschlecht ♂/♀	17/19	5/7	5/6	18/16	35/47	24/36
Alter M (SD)*	24,9 (2,3)	35,8 (3,0)	46,3 (3,0)	57,0 (2,8)	65,3 (2,9)	74,3 (2,5)
Bildung M (SD)*	16,1 (1,9)	15,7 (2,3)	14,6 (2,1)	14,7 (2,6)	15,0 (2,6)	14,6 (2,5)

\*Angaben in Jahren

Für den Parameter Alter sind die Voraussetzungen der Varianzanalyse (ANOVA), Normalverteilung und Homoskedastizität der Residuen, erfüllt (Kolmogorov-Smirnov-Test:  $z=1,184$ ,  $p=.121$ , Levene-Test:  $F(5,229)=1,540$ ,  $p=.178$ ). Naturgemäß unterscheiden sich die Altersgruppen in der Varianzanalyse hinsichtlich des Alters ( $F(5,229)=1812,700$ ,  $p=.000$ ). Post-hoc-t-Tests für unabhängige Stichproben mit Bonferroni-Holm-Adjustierung verweisen dabei auf signifikante Unterschiede zwischen den benachbart liegenden Gruppen (Testgrößen siehe Anhang 5.3., alle  $p=.000$ ).

Für den Parameter Bildung ist die Annahme der Homoskedastizität verletzt (Levene-Test:  $F(2,229)=3,277$ ,  $p=.007$ ), so dass die Varianzanalyse nicht durchgeführt werden konnte. In der Kruskal-Wallis-Rangvarianzanalyse unterscheiden sich die Gruppen nicht signifikant ( $\chi^2(2)=10,733$ ,  $p=.057$ ). Der Chi-Quadrat-Test indiziert zudem keine Unterschiede zwischen der Geschlechterverteilung innerhalb der sechs Altersgruppen ( $\chi^2(5)=1,730$ ,  $p=.885$ ).

*Teilnehmer in der klinischen Studie*

In Tabelle 4 sind die demografischen Charakteristika der Patienten mit leichter kognitiver Störung (MCI) bzw. Demenz bei Alzheimerkrankheit (DAT) denen der gesunden Kontrollgruppe (KG) des gleichen Altersbereichs (65–80 Jahre) gegenübergestellt.

**Tabelle 4: Demografische Eigenschaften der Diagnosegruppen der klinischen Studie**

	Diagnosegruppe*		
	KG (n = 107)	MCI (n = 82)	DAT (n = 36)
Geschlecht ♂/♀	48/59	48/34	20/16
Alter M (SD)**	71,3 (4,1)	73,6 (4,1)	73,9 (3,7)
Bildung M (SD)**	14,6 (2,6)	13,9 (2,7)	13,8 (2,6)

\* Kontrollgruppe (KG), Patienten mit leichter kognitiver Störung (MCI) und Demenz bei Alzheimerkrankheit (DAT)

\*\* Angaben in Jahren

Für den Parameter Alter sind die Voraussetzungen der ANOVA, Normalverteilung und Homoskedastizität der Residuen, erfüllt (Kolmogorov-Smirnov-Test:  $z=1,146$ ,  $p=.144$ ; Levene-Test:  $F(2,222)=0,883$ ,  $p=.415$ ). Die ANOVA verweist auf signifikante Altersunterschiede zwischen den Gruppen ( $F(2,222)=9,809$ ,  $p=.000$ ). Post-hoc-t-Tests für unabhängige Stichproben mit Bonferroni-Holm-Adjustierung belegen dabei ein signifikant geringeres Alter der Kontrollgruppe im Vergleich zu beiden Patientengruppen ( $t(187)_{KG,MCI}=-3,771$ ,  $p=.002$ ;  $t(141)_{KG,DAT}=-3,406$ ,  $p=.003$ ), jedoch keinen Unterschied zwischen den beiden Patientengruppen ( $t(116)_{MCI,DAT}=-0,448$ ,  $p=.655$ ).

Für den Parameter Bildung ist die Annahme einer Normalverteilung der Residuen verletzt (Kolmogorov-Smirnov-Test:  $z=2,062$ ,  $p=.000$ ), so dass die Varianzanalyse nicht durchgeführt werden kann. Die Untersuchungsgruppen unterscheiden sich im Kruskal-Wallis-Test nicht in Bezug auf die Bildungsdauer ( $\chi^2(2)=4,479$ ,  $p=.107$ ).

Hinsichtlich der Geschlechterverteilung unterscheiden sich die Gruppen im Chi-Quadrat-Test ebenfalls nicht ( $\chi^2(2)=3,751$ ,  $p=.153$ ).

### 3.1.2. Testleistungen in der Stichprobe der gesunden Probanden

Die statistischen Kennwerte der Bearbeitungszeiten im Berliner Shape Trail Test (B-STT) sind in Tabelle 5 getrennt nach Altersgruppen dargestellt. Zum Vergleich mit aufgeführt sind die entsprechenden Kennwerte der Bearbeitungszeiten im Trail Making Test A (TMT-A) und Trail Making Test B (TMT-B). Angegeben sind jeweils die Anzahl der untersuchten Personen (n), Mittelwert (M), Median (MD), Standardabweichung (SD), Minimum (Min) und Maximum (Max).

**Tabelle 5: Statistische Kennwerte der Bearbeitungszeiten im B-STT, TMT-A und TMT-B bei gesunden Probanden**

Altersgruppe	Berliner Shape Trail Test (Bearbeitungszeit <sup>*</sup> )					
	n	M	MD	SD	Min	Max
21–30	36	30,1	30,5	5,8	17	43
31–40	12	33,9	31,0	9,5	23	54
41–50	11	41,5	46,0	9,1	28	59
51–60	34	44,5	40,0	13,9	25	86
61–70	82	51,7	48,5	14,9	25	99
71–80	60	63,5	59,5	19,7	34	126
Altersgruppe	Trail Making Test A (Bearbeitungszeit <sup>*</sup> )					
	n	M	MD	SD	Min	Max
21–30	36	24,8	24,0	5,7	17	37
31–40	12	28,4	27,5	8,3	17	44
41–50	11	35,9	31,0	14,6	24	77
51–60	34	34,6	33,5	8,6	21	59
61–70	82	38,7	37,0	9,7	18	69
71–80	60	46,4	44,5	13,3	26	101
Altersgruppe	Trail Making Test B (Bearbeitungszeit <sup>*</sup> )					
	n	M	MD	SD	Min	Max
21–30	36	49,8	47,5	14,1	29	94
31–40	12	60,8	57,0	19,2	28	102
41–50	11	62,6	62,0	16,9	40	91
51–60	34	76,1	67,5	25,0	42	154
61–70	82	80,1	77,0	20,4	41	134
71–80	60	98,8	91,5	30,6	44	182

\* Angaben in Sekunden

### 3.1.3. Testleistungen in der klinischen Studie

Die statistischen Kennwerte der Bearbeitungszeiten im B-STT sind für die Patienten mit MCI und die Patienten mit DAT sowie für die gesunde Kontrollgruppe im Alter von 65–80 Jahren (KG) in Tabelle 6 gegenübergestellt.

**Tabelle 6: Statistische Kennwerte der Bearbeitungszeiten im B-STT in der klinischen Studie**

Gruppe	Berliner Shape Trail Test (Bearbeitungszeit <sup>*</sup> )					
	n	M	MD	SD	Min	Max
KG	107	58,8	56,0	18,3	25	126
MCI	82	92,8	82,0	43,2	35	246
DAT	36	149,2	138,0	70,8	49	319

\* Angaben in Sekunden

### 3.1.4. Neuropsychologische Charakteristika in der klinischen Studie

Die Rohwerte der Untertests der neuropsychologischen Batterie CERAD-Plus einschließlich des Trail Making Tests sind für die Patienten mit MCI und DAT sowie für die gesunde Kontrollgruppe im Altersbereich von 65–80 Jahren in Tabelle 7 aufgeführt (Mittelwert, Standardabweichung). Angegeben sind weiterhin auf lokalem Signifikanzniveau vorliegende Unterschiede zwischen den Gruppen. Diese basieren auf der ANOVA und nachfolgenden t-Tests für unabhängige Stichproben sowie bei Verletzung der Voraussetzungen der ANOVA (Prüfgrößen siehe Anhang 5.4.) auf den Ergebnissen des Kruskal-Wallis-Tests und den Paarvergleichen mit dem Wilcoxon-Mann-Whitney-Test.

Die Parameter der CERAD-Plus sind in der Tabelle wie folgt bezeichnet:

- MMST: Summenscore im Mini-Mental-Status-Test
- WL<sub>Ges</sub>: Summe der in den drei Lerndurchgängen der Wortliste genannten Wörter
- WL<sub>Abr</sub>: Zahl der in der verzögerten Abrufbedingung der Wortlisten genannten Wörter
- WL<sub>Rek</sub>: Summe der richtig als bekannt bzw. neu erkannten Wörter der Wortliste
- BNT: Summe der richtig erkannten und benannten Objekte des Boston Naming Tests
- VKon: Punktwert in der Visuokonstruktionsaufgabe
- FigG: Punktwert im verzögerten Abruf der Figuren
- Flu-T / Flu-S: Anzahl der genannten Tiere / S-Wörter in den Wortflüssigkeitsaufgaben
- TMT-A / TMT-B: Trail Making Test A / Trail Making Test B

Leistungsunterschiede zwischen den Gruppen wurden in allen Testverfahren mit Ausnahme der Visuokonstruktionsaufgabe nachgewiesen. Der Paarvergleich wies darüber hinaus nach, dass in allen Testverfahren die höchsten Leistungen von der Kontrollgruppe erbracht wurden. Zudem waren die Leistungen der Patientengruppe mit MCI in allen Verfahren mit Ausnahme der Wortflüssigkeit für S-Wörter höher als die der Patientengruppe mit DAT. Aus Gründen der Übersichtlichkeit wurde in der Tabelle auf eine genaue Darstellung der Prüfgrößen und Signifikanzwerte der Testverfahren verzichtet; diese befindet sich in Anhang 5.5.

**Tabelle 7: Neuropsychologische Charakteristika in der klinischen Studie**

Test		KG (n = 107)	MCI (n = 82)	DAT (n = 36)	Signifikante Kontraste (p<.05)
MMST	M (SD)	28,6 (1,2)	27,5 (1,8)	24,6 (2,5)	KG>MCI>DAT
WL <sub>Ges</sub>	M (SD)	22,5 (3,2)*	20,4 (3,8)	14,1 (3,3)	KG>MCI>DAT
WL <sub>Abr</sub>	M (SD)	7,9 (16)*	6,1 (2,2)	2,4 (2,0)	KG>MCI>DAT
WL <sub>Rek</sub>	M (SD)	19,8 (0,5)*	19,4 (1,1)	17,0 (2,5)	KG>MCI>DAT
BNT	M (SD)	14,7 (0,6)*	14,3 (1,1)	13,6 (2,0)	KG>MCI>DAT
VKon	M (SD)	10,0 (0,9)*	10,0 (1,0)	9,8 (1,0)	nicht signifikant
FigG	M (SD)	9,0 (1,4)*	7,1 (2,6)	3,7 (2,8)	KG>MCI>DAT
Flu-T	M (SD)	23,9 (5,4)*	19,9 (6,3)	14,9 (4,9)	KG>MCI>DAT
Flu-S	M (SD)	15,6 (4,3)*	12,1 (4,0)**	10,8 (4,1)**	KG>MCI, KG>DAT
TMT-A	M (SD)	43,4 (12,5)	60,1 (23,8)	71,5 (30,0)	KG<MCI<DAT
TMT-B	M (SD)	92,3 (27,9)	159,9 (74,3)	242,2 (132,2)	KG<MCI<DAT

\* Bei KG absolvierten nur 97 Personen die gesamte Testbatterie.

\*\* Bei MCI und DAT wurde in jeweils einem Fall der Untertest S-Wörter nicht durchgeführt.

### **3.2. Einfluss demografischer Faktoren auf die Bearbeitungszeit im B-STT**

#### **3.2.1. Einfluss von Alter und Bildung (korrelative Untersuchungen)**

Der Pearson-Korrelationskoeffizient verweist auf einen signifikanten, stark ausgeprägten Zusammenhang zwischen dem Alter der Probanden und der Bearbeitungszeit im B-STT ( $r(233)=0,606$ ,  $p=.000$ ). Der Anteil der durch das Alter aufgeklärten Varianz an der Gesamtvarianz der Bearbeitungszeit liegt bei 36,7 %. Ein höheres Alter steht dabei mit einer langsameren Bearbeitung der Aufgabe in Verbindung. Hinsichtlich des Einflusses der Bildungsdauer auf die Testleistungen offenbaren sich signifikante, jedoch schwache Assoziationen ( $r(233)=-0,210$ ,  $p=.004$ ). Der Anteil der Varianzaufklärung beträgt 4,4 %. Eine höhere Bildung ist dabei mit einer rascheren Bearbeitung des B-STT verbunden.

### 3.2.2. Einfluss von Alter und Bildung (multiple Regression)

Vor Durchführung bzw. Bewertung der Regressionsanalyse werden ihre statistischen Voraussetzungen geprüft:

1. Normalverteilung der Residuen: Der Kolmogorov-Smirnov-Test belegt, dass die Verteilung der Residuen einer Normalverteilung entspricht ( $z=0,622$ ,  $p=.834$ ).
2. Homoskedastizität der Residuen: Der Levene-Test indiziert das Vorliegen homogener Fehlervarianzen zwischen den beiden anhand der Bildungsränge gebildeten Gruppen ( $F(1,233)=0,595$ ,  $p=.441$ ).
3. Keine Autokorrelation der Residuen: Die Durbin-Watson-Statistik zeigt einen Wert von 2,037 an. Es liegt damit keine Autokorrelation der Residuen vor.
4. Keine Multikollinearität: Die Korrelation zwischen den beiden unabhängigen Variablen Alter und Bildung ist signifikant, aber sehr schwach ausgeprägt ( $r(233)=-.181$ ,  $p=.005$ ). Der in der Regressionsanalyse ermittelte Varianzinflationsfaktor liegt bei 1,034 und weist damit ebenfalls auf ein Fehlen von Multikollinearität hin.

Das Regressionsmodell ist statistisch signifikant ( $F(2,234)=102,494$ ,  $p=.000$ ). Ein konstanter Faktor und die beiden Prädiktoren sind signifikant an dem Modell beteiligt ( $t_{\text{Konstante}}(233)=24,359$ ,  $p=.000$ ;  $t_{\text{Alter}}(233)=13,588$ ,  $p=.000$ ;  $t_{\text{Bildung}}(233)=-1,980$ ,  $p=.049$ ).

Die multiple Korrelation der Prädiktoren mit dem Kriterium liegt bei  $r=.685$ . Dementsprechend liegt der Anteil der durch das Modell erklärten Varianz an der Gesamtvarianz der Bearbeitungszeiten bei 46,9 %. Die Regressionsgleichung im ausgegebenen Modell lautet:

$$\text{Log}_{10}(\text{Bearbeitungszeit}) = 1,404 + 0,006 \times \text{Alter} - 0,006 \times \text{Bildungsdauer}$$

### 3.2.3. Einfluss des Geschlechts

Vor Durchführung bzw. Bewertung der Kovarianzanalyse (ANCOVA) wurden ihre statistischen Voraussetzungen geprüft:

1. Normalverteilung der Residuen: Der Kolmogorov-Smirnov-Test belegt, dass die Verteilung der Residuen einer Normalverteilung entspricht ( $z=0,471$ ,  $p=.980$ ).
2. Homoskedastizität der Residuen: Der Levene-Test indiziert das Vorliegen homogener Fehlervarianzen zwischen den Gruppen ( $F(1,233)=0,306$ ,  $p=.581$ ).
3. Korrelation der Kovariaten mit der abhängigen Variablen: Siehe Abschnitt 3.2.1. Die Bedingung ist erfüllt.
4. Homogenität der Regressionsgeraden: Keiner der mit der ANCOVA ermittelten Interaktionseffekte zwischen dem Geschlecht (G) der Probanden und dem Alter (A), dem Bildungsniveau (B) bzw. beiden Variablen erreicht Signifikanzniveau ( $F_{GA}(1,233)=1,167$ ,  $p=.281$ ;  $F_{GB}(1,233)=0,677$ ,  $p=.411$ ;  $F_{GAB}(1,233)=1,193$ ,  $p=.276$ ). Von einer Homogenität der Regressionsgeraden ist daher auszugehen.

Die Kovarianzanalyse zeigt, dass das Geschlecht der Probanden, bei Kontrolle von Bildungsniveau und Alter, keinen signifikanten Einfluss auf die Bearbeitungszeit im B-STT hat ( $F(1,233)=0,353$ ,  $p=.553$ ).

### 3.3. Test-Retest-Reliabilität des B-STT

Tabelle 8 gibt die Datenbasis für die Analyse der zeitlichen Stabilität der Bearbeitungszeit im B-STT wieder.

**Tabelle 8: Zeitverlauf der Bearbeitungszeiten (in Sekunden) im B-STT**

Messzeitpunkt	n	M	MD	SD	Min	Max
T1	19	51,2	47,0	19,5	25	97
T2	19	48,0	46,0	18,0	27	107

Die Daten beider Messzeitpunkte (T1, T2) sind normalverteilt (Kolmogorov-Smirnov-Test:  $z_{T1}=0,605$ ,  $p=.857$ ;  $z_{T2}=0,974$ ,  $p=.299$ ). Die Reliabilitätsprüfung des B-STT zeigt einen signifikanten, starken Zusammenhang zwischen den Bearbeitungszeiten der beiden

Messzeitpunkte ( $r(17)=.793$ ,  $p=.000$ ). 62,9 % der Gesamtvarianz der Bearbeitungszeit eines Messzeitpunktes wird durch die Bearbeitungszeit zum anderen Messzeitpunkt aufgeklärt.

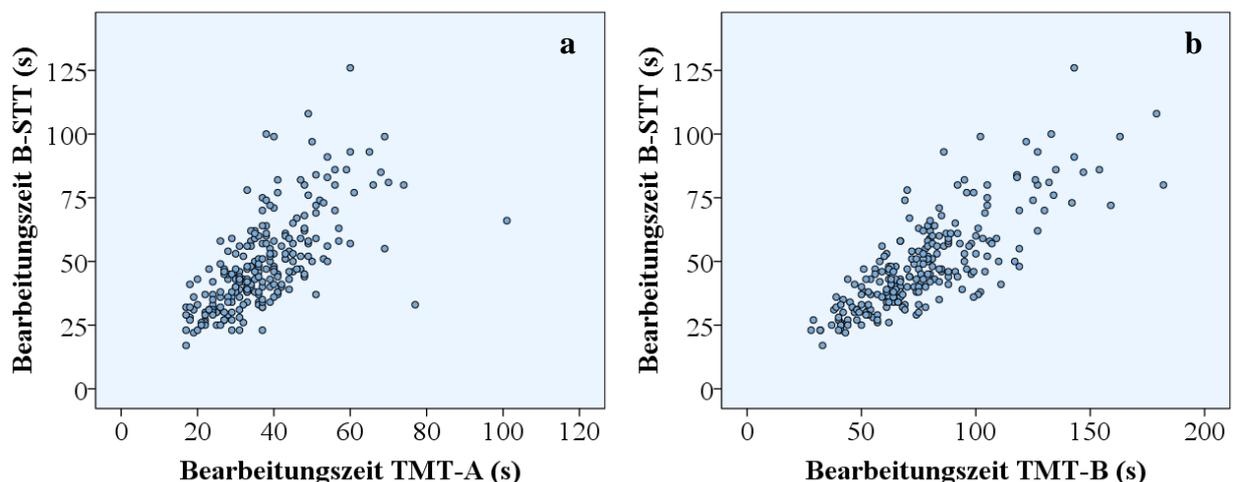
### 3.4. Validität des B-STT

#### 3.4.1. Konstruktvalidität

##### *Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT bei gesunden Probanden*

Die Korrelationen zwischen den Bearbeitungszeiten im B-STT und TMT-A bzw. B-STT und TMT-B erreichen Signifikanzniveau (beide  $p=.000$ ). Der Zusammenhang zwischen dem B-STT und dem TMT-A ist dabei stark ( $r_{SA}(233)=.673$ ), der Zusammenhang zwischen dem B-STT und dem TMT-B sehr stark ausgeprägt ( $r_{SB}(233)=.803$ ). Die Zusammenhänge sind positiv: Eine längere Bearbeitungszeit in einem Verfahren steht mit einer längeren Bearbeitungszeit im anderen Verfahren in Verbindung. Der Anteil der aufgeklärten Varianz an der Gesamtvarianz des jeweils anderen Verfahrens liegt bei 45,3 % für das Paar B-STT/TMT-A und bei 64,4 % für das Paar B-STT/TMT-B. Der Vergleich der Korrelationskoeffizienten  $r_{SA}$  und  $r_{SB}$  zeigt eine signifikant höhere Stärke des Zusammenhangs zwischen den Bearbeitungszeiten von B-STT und TMT-B ( $z=3,317$ ,  $p=.000$ ).

Zur Veranschaulichung sind die Zusammenhänge zwischen den Bearbeitungszeiten im B-STT und im TMT-A bzw. TMT-B in Abbildung 3 grafisch dargestellt.

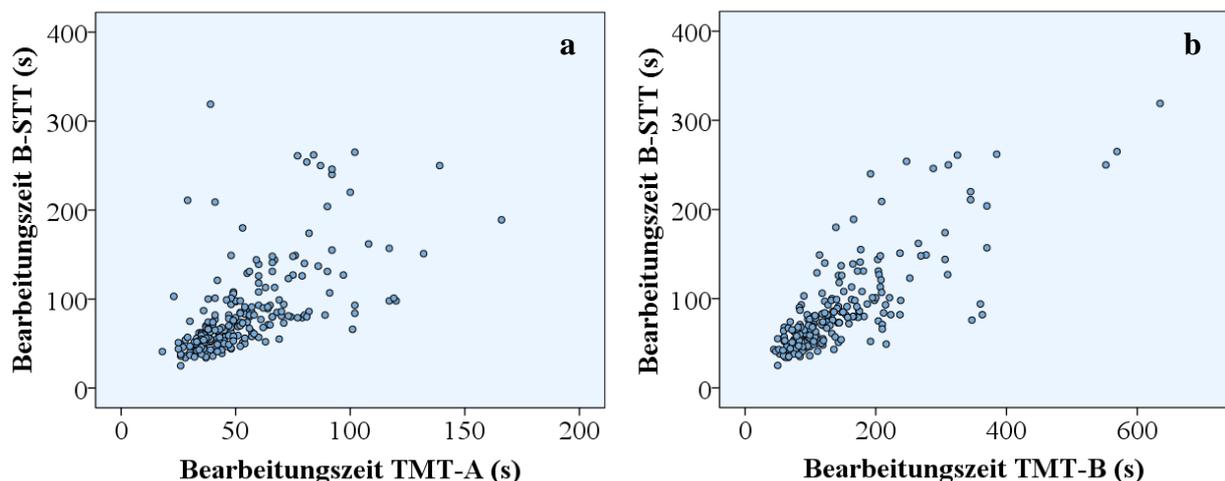


**Abbildung 3: Darstellung der Beziehung zwischen den Bearbeitungszeiten in B-STT und TMT-A (a) bzw. B-STT und TMT-B (b) in der Stichprobe der gesunden Probanden**

### Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT in der klinischen Studie

Die folgende Zusammenhanganalyse bezieht sich auf die Gesamtstichprobe der 65- bis 80-jährigen Studienteilnehmer. Diese beinhaltet Patienten mit MCI und DAT sowie alle gesunden Probanden der Altersgruppe. Die Korrelationen zwischen den Bearbeitungszeiten im B-STT und TMT-A bzw. B-STT und TMT-B erreichen Signifikanzniveau (beide  $p=.000$ ); die Testleistungen sind jeweils positiv assoziiert. Der Zusammenhang zwischen B-STT und TMT-A ist stark ( $r_{SA}(223)=.604$ ), der Zusammenhang zwischen B-STT und TMT-B sehr stark ( $r_{SB}(223)=.802$ ) ausgeprägt. Die Varianzaufklärung beträgt somit 36,5 % (B-STT/TMT-A) bzw. 64,3 % (B-STT/TMT-B). Der Vergleich der Korrelationskoeffizienten  $r_{SA}$  und  $r_{SB}$  zeigt eine signifikant höhere Stärke des Zusammenhangs zwischen dem B-STT und dem TMT-B ( $z=5,122$ ,  $p=.000$ ).

Zur Veranschaulichung sind die genannten Zusammenhänge in Abbildung 4 grafisch dargestellt.



**Abbildung 4:** Darstellung der Beziehung zwischen den Bearbeitungszeiten in B-STT und TMT-A (a) bzw. B-STT und TMT-B (b) in der klinischen Studie

### Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT in der klinischen Studie bei Kontrolle der globalen kognitiven Funktion

Die Zusammenhänge zwischen den Bearbeitungszeiten in den Testverfahren bei Ausparialisierung des Summenwertes im MMST stellen sich wie folgt dar: Ein mittlerer Zusammenhang findet sich zwischen B-STT und TMT-A ( $r_{SA,M}(222)=.543$ ,  $p=.000$ ). Ein starker Zusammenhang findet sich zwischen dem B-STT und dem TMT-B ( $r_{SB,M}(222)=.715$ ,  $p=.000$ ). Die Varianzaufklärung beträgt somit 29,5 % (B-STT/TMT-A) bzw. 51,1 % (B-STT/TMT-B). Der Zusammenhang zwischen B-STT und TMT-B ist dabei auch unter Berücksichtigung des globalen kognitiven Funktionsniveaus stärker ausgeprägt ( $z=3,579$ ,  $p=.000$ ).

*Zusammenhang zwischen den Bearbeitungszeiten im B-STT und TMT-B und anderen neuropsychologischen Verfahren*

Die Stärke der Zusammenhänge zwischen den Bearbeitungszeiten von B-STT und TMT-B und den neuropsychologischen Tests der CERAD-Plus in der Gesamtpopulation der 65- bis 80-Jährigen (KG, MCI, DAT) kann Tabelle 9 entnommen werden (Bezeichnungen siehe Abschnitt 3.1.4.). Angegeben sind die Korrelationskoeffizienten und lokalen Signifikanzwerte nach Eliminierung des MMST-Gesamtscores als partielle Korrelation. Auf eine Adjustierung des Signifikanzniveaus wurde verzichtet, da eine deskriptive Betrachtung der Korrelationskoeffizienten als ausreichend erachtet wurde.

Die Assoziationen der Leistungsparameter sind insgesamt als sehr schwach bis schwach zu bewerten. Der auf deskriptiver Ebene höchste Zusammenhang, formal schwach ausgeprägt, besteht zwischen der Bearbeitungszeit im B-STT und der Wortflüssigkeit für Tiere (Varianzaufklärung 10,8 %). Alle auf lokalem Niveau signifikanten Zusammenhänge sind invers ausgeprägt, eine schnellere Bearbeitung bzw. weniger Fehler stehen mit höheren Testwerten der CERAD-Plus in Verbindung.

**Tabelle 9: Partialkorrelationen zwischen den Bearbeitungszeiten im B-STT und TMT-B und den weiteren Leistungsparametern der CERAD-Plus\* bei allen 65- bis 80-jährigen Studienteilnehmern\*\***

CERAD-Plus	B-STT Zeit	TMT-B Zeit
WL <sub>Ges</sub>	$r(212)=-.254, p=.000$	$r(212)=-.291, p=.000$
WL <sub>Abr</sub>	$r(212)=-.167, p=.015$	$r(212)=-.174, p=.010$
WL <sub>Rek</sub>	$r(212)=-.196, p=.004$	$r(212)=-.240, p=.000$
BNT	$r(212)=-.115, p=.093$	$r(212)=-.068, p=.320$
VKon	$r(212)=-.205, p=.003$	$r(212)=-.162, p=.017$
FigG	$r(212)=-.261, p=.000$	$r(210)=-.198, p=.004$
Flu-T	$r(212)=-.329, p=.000$	$r(212)=-.262, p=.000$
Flu-S	$r(210)=-.275, p=.000$	$r(210)=-.274, p=.000$

\* MMST-Gesamtscore auspartialisiert

\*\* gesunde Kontrollgruppe, Patienten mit MCI und Patienten mit DAT

### 3.4.2. Kriteriumsvalidität

#### *Leistungsunterschiede der Gruppen hinsichtlich der Bearbeitungszeit im B-STT*

Die Durchführung einer ANOVA für die Gruppenunterschiede der Bearbeitungszeiten konnte aufgrund der Verletzung der Homoskedastizität der Residuen (Levene-Test:  $F(2,222)=7,444$ ,  $p=.001$ ) und bei deutlichen Unterschieden der Gruppengrößen der Stichproben nicht erfolgen.

Der Kruskal-Wallis-Test zeigt signifikante Unterschiede zwischen den Bearbeitungszeiten der Gruppen ( $\chi^2(2)=87,971$ ,  $p=.000$ ). Die Post-Hoc-Vergleiche mit dem Wilcoxon-Mann-Whitney-Test indizieren signifikante Unterschiede zwischen allen Paaren ( $z_{KG.MCI}=-6,895$ ;  $z_{KG.DAT}=-7,748$ ;  $z_{MCI.DAT}=-4,553$ ; alle  $p=.000$ ). Die gesunde Kontrollgruppe zeigt dabei die besten Leistungen, die Patienten mit DAT die am stärksten beeinträchtigten Leistungen.

#### *Effektstärken der Gruppenunterschiede zwischen KG, MCI und DAT*

Die Effektstärken Cohens  $d$  der Gruppenunterschiede sind für den B-STT in Tabelle 10 wiedergegeben. Zum Vergleich sind auch die Effektstärken der Maße des TMT-B aufgeführt. Bezüglich der Bearbeitungszeiten indizieren sie starke Effekte im Vergleich der Kontrollgruppe mit den klinischen Populationen in beiden Testverfahren. So unterscheiden sich die Bearbeitungszeiten zwischen Patienten mit DAT und der gesunden Kontrollgruppe um mehr als 1,5 Standardabweichungen. Der Unterschied zwischen den beiden klinischen Gruppen ist im B-STT stark, im TMT-B mittel bis stark ausgeprägt.

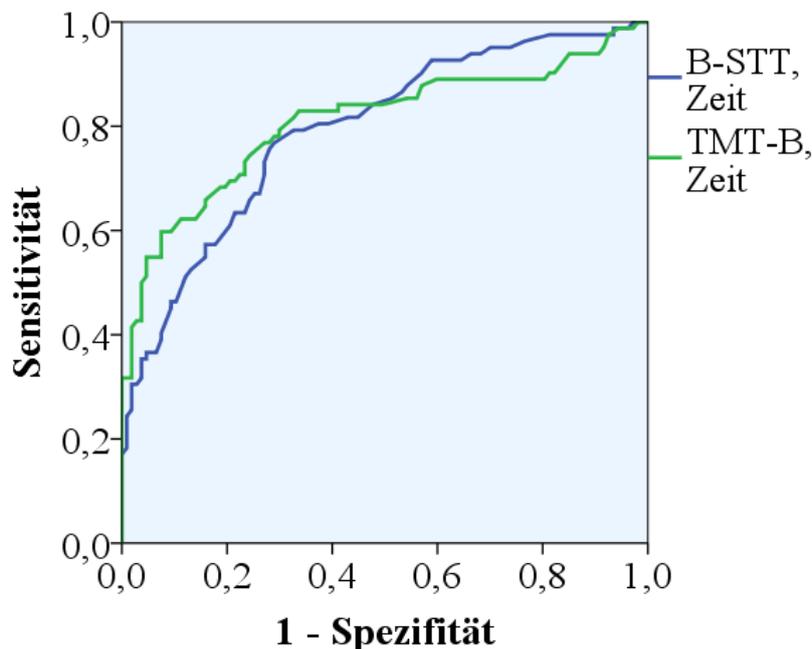
**Tabelle 10: Effektstärken der Gruppenunterschiede in den Bearbeitungszeiten von B-STT und TMT-B als Cohens  $d$**

Gruppenkontrast	Effektstärke des Gruppenunterschieds der Bearbeitungszeiten		
	KG vs. MCI	KG vs. DAT	MCI vs. DAT
B-STT Zeit	$d=1,13$	$d=1,70$	$d=0,92$
TMT-B Zeit	$d=1,21$	$d=1,83$	$d=0,70$

### Trennschärfe des B-STT

Abbildung 5 zeigt die *Receiver-Operating-Characteristic-Kurve* (ROC-Kurve) für die Bearbeitungszeit im B-STT bei der Differenzierung zwischen der Kontrollgruppe und Patienten mit leichter kognitiver Störung (KG vs. MCI). Die entsprechende ROC-Kurve des TMT-B ist zum Vergleich ebenfalls abgebildet. Die Fläche unter der Kurve (*area under the curve*; AUC) ist dabei Maß für die Trennschärfe der Verfahren.

Die AUC hat für den B-STT einen Wert von 0,79 (95 %-Konfidenzintervall: 0,73–0,86) und für den TMT-B einen Wert von 0,81 (95 %-Konfidenzintervall: 0,74–0,88). Nach den Konventionen von Swets [106] sind die Trennschärfen der Bearbeitungszeiten jeweils als angemessen einzuordnen.



**Abbildung 5: ROC-Kurven für die Bearbeitungszeiten von B-STT und TMT-B bei der Differenzierung zwischen der Kontrollgruppe und Patienten mit MCI**

Die ROC-Kurven für die Differenzierung zwischen der Kontrollgruppe und der Gruppe der Patienten mit Demenz bei Alzheimerkrankheit (KG vs. DAT) sind in Abbildung 6 dargestellt. Für die Bearbeitungszeit erreicht die AUC beim B-STT einen Wert von 0,93 (95 %-Konfidenzintervall: 0,88–0,98) und beim TMT-B einen Wert von 0,95 (95 %-Konfidenzintervall: 0,91–0,99). Die Trennschärfen der Bearbeitungszeiten liegen damit im sehr guten Bereich.

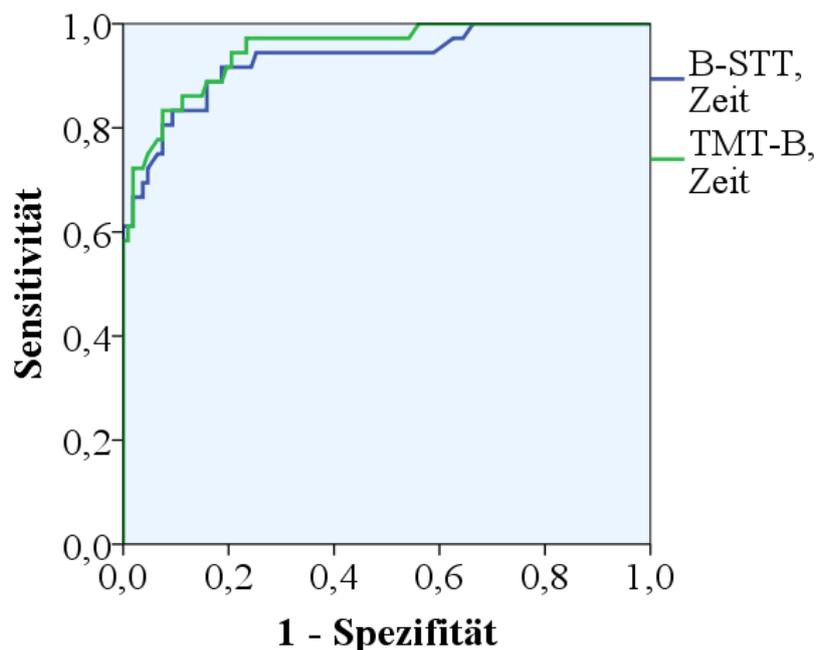


Abbildung 6: ROC-Kurven für die Bearbeitungszeiten von B-STT und TMT-B bei der Differenzierung zwischen der gesunden Kontrollgruppe und Patienten mit DAT

Die ROC-Kurven für die Detektion einer Demenz innerhalb der Gruppe der Patienten (MCI vs. DAT) sind in Abbildung 7 dargestellt. Für die Bearbeitungszeit beträgt die AUC beim B-STT 0,76 (95 %-Konfidenzintervall: 0,69–0,86). Beim TMT-B beträgt sie 0,71 (95 %-Konfidenzintervall: 0,61–0,81). Die Trennschärfen sind als angemessen zu bewerten.

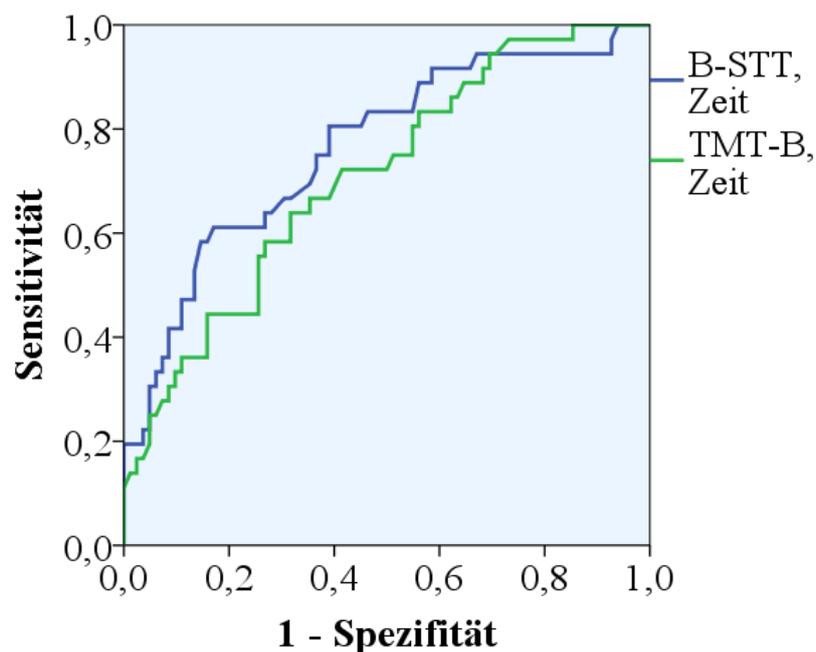


Abbildung 7: ROC-Kurven für die Bearbeitungszeiten von B-STT und TMT-B für die Detektion einer Demenz in der Patientengruppe (DAT vs. MCI)

### 3.5. Unterschiede in den Bearbeitungszeiten von B-STT und TMT-B

#### 3.5.1. Gruppenübergreifende Prüfung

Die inferentielle Analyse erfolgte zunächst über alle Gruppen (KG, MCI, DAT) hinweg. Ein t-Test für unabhängige Stichproben konnte aufgrund der Verletzung der Annahme einer Normalverteilung der Differenzwerte (Kolmogorov-Smirnov-Test:  $z=3,537$ ,  $p=.000$ ) nicht durchgeführt werden. Der Wilcoxon-Vorzeichen-Rang-Test verweist auf eine signifikant schnellere Bearbeitung des B-STT im Vergleich zum TMT-B ( $z = -15,655$ ,  $p=.000$ ) in der Gesamtstichprobe der gesunden Probanden und der Patienten. Es handelt sich um einen starken Effekt: Die Effektgröße des Unterschieds liegt bei  $d=3,01$ .

#### 3.5.2. Vergleich der Untersuchungsgruppen

In Tabelle 11 sind für die Probanden im Alter von unter 65 bzw. ab 65 Jahren sowie für die beiden klinischen Gruppen die mittleren Differenzwerte der Bearbeitungszeiten von TMT-B und B-STT mit Standardabweichungen angegeben. Die positiven Differenzwerte spiegeln dabei die im Mittel raschere Bearbeitung des B-STT wider. Weiterhin ist die Zahl der Personen angegeben, die für die Bearbeitung des B-STT weniger Zeit benötigte als für die Bearbeitung des TMT-B.

**Tabelle 11: Differenz der Bearbeitungszeiten von TMT-B und B-STT und Anzahl der Personen, die den B-STT schneller absolvierten als den TMT-B getrennt nach Untersuchungsgruppen**

	Probanden 21–64 Jahre n = 128	Probanden 65–80 Jahre n = 107	MCI n = 82	DAT n = 36
TMT-B–BSTT* M (SD)	24,9 (14,7)	33,6 (18,9)	67,0 (59,4)	93,0 (92,0)
$n_{\text{TMTB}>\text{BSTT}}$	127	105	72	31

\* Angabe in Sekunden

Eine Varianzanalyse zum Vergleich der Differenzwerte konnte aufgrund der Verletzung der Annahmen einer Normalverteilung und Homoskedastizität der Residuen nicht durchgeführt werden (Kolmogorov-Smirnov-Test  $z=2,767$ ,  $p=.000$ ; Levene-Test:  $F(3,349)=57,349$ ,  $p=.000$ ).

Der Kruskal-Wallis-Test offenbart signifikante Unterschiede zwischen den vier Untersuchungsgruppen (Probanden [21–64 Jahre], Probanden [65–80 Jahre], MCI, DAT) in Bezug auf die Differenz zwischen den Bearbeitungszeiten von TMT-B und B-STT ( $\chi^2(3)=64,593$ ,  $p=.000$ ).

Die Post-hoc-Vergleiche mit dem Wilcoxon-Mann-Whitney-Test zeigen signifikante Unterschiede zwischen allen Untersuchungsgruppen mit Ausnahme des Kontrasts der beiden Patientengruppen ( $z_{\text{Jung/Alt}}=-3,778$ ,  $p=.000$ ;  $z_{\text{Jung/MCI}}=-6,541$ ,  $p=.000$ ;  $z_{\text{KGJung/DAT}}=-4,888$ ,  $p=.000$ ;  $z_{\text{Alt/MCI}}=-4,851$ ,  $p=.000$ ;  $z_{\text{Alt/DAT}}=-4,082$ ,  $p=.000$ ;  $z_{\text{MCI/DAT}}=-1,298$ ,  $p=.194$ ). Die Zeitersparnis ist dementsprechend bei älteren Probanden stärker ausgeprägt als bei jüngeren Probanden und bei Patienten mit kognitiven Beeinträchtigungen höher als bei gesunden Personen.

---

## 4. Diskussion

### 4.1. Ziel der Untersuchung

Die vorliegende Arbeit stellt die Ersterprobung des Berliner Shape Trail Tests (B-STT) dar, eines neuen, als Alternative zum Trail Making Test B entwickelten neuropsychologischen Testverfahrens zur Erfassung exekutiver Funktionen. Der B-STT wurde im Rahmen der Untersuchung an insgesamt 235 gesunden Personen verschiedener Altersgruppen sowie an 82 Patienten mit leichter kognitiver Störung (MCI) und 36 Patienten mit Demenz bei Alzheimerkrankheit (DAT) eingesetzt. Hauptziele der Untersuchung waren, die Eignung des B-STT als neuropsychologisches Instrument und die Vergleichbarkeit des B-STT mit dem Trail Making Test B (TMT-B), einem bereits etablierten Verfahren, zu untersuchen. Neben einer Charakterisierung der demografischen Eigenschaften des B-STT wurde dazu besonderes Augenmerk auf die Konstrukt- und Kriteriumsvalidität gelegt. Weiterhin wurden die Testgütekriterien Reliabilität und Ökonomie untersucht.

### 4.2. Zusammenfassung der Ergebnisse

In den folgenden Abschnitten werden die Ergebnisse der Studie zusammengefasst und im Kontext der übergeordneten Fragestellungen und konkreten Forschungshypothesen diskutiert. Eine weiterführende Diskussion und Bewertung der Befunde in der Zusammenschau mit den in Abschnitt 4.3. vorgestellten Limitationen der Studie sowie der aktuellen Literatur folgt in den Abschnitten 4.4. bis 4.6.

#### 4.2.1. Zusammenhang zwischen den demografischen Merkmalen der Probanden und der Bearbeitungszeit im B-STT

Es besteht ein stark ausgeprägter Zusammenhang zwischen dem Alter der Probanden und der Bearbeitungszeit beim B-STT. Dieser Zusammenhang ist positiv gerichtet: Mit zunehmendem Alter benötigen die Probanden mehr Zeit. Der Zusammenhang zwischen der Bildungsdauer der Probanden und ihren Bearbeitungszeiten ist ebenfalls signifikant, jedoch nur schwach ausgeprägt. Ein höheres Bildungsniveau ist dabei mit niedrigeren Bearbeitungszeiten assoziiert. Die in Abschnitt 1.9. formulierte Hypothese kann somit angenommen werden.

Die Ergebnisse der multiplen Regressionsanalyse, und hier insbesondere die multiple Korrelation der Prädiktoren mit dem Kriterium, deuten zudem auf deskriptiver Ebene darauf hin, dass eine

---

Kombination von Alter und Bildung einen größeren Teil der Varianz der Testleistungen erklären kann als beide Faktoren für sich.

Die Bearbeitungszeiten und Fehlerzahlen von Männern und Frauen unterscheiden sich nicht signifikant; das Geschlecht kann also bei der Interpretation der Testleistungen vernachlässigt werden.

#### *4.2.2. Test-Retest-Reliabilität des B-STT*

Die in der Untersuchung ermittelte Test-Retest-Reliabilität liegt bei  $r=.799$ , was einem starken bis sehr starken Zusammenhang der Leistungen entspricht. Die Stichprobe ist ausreichend groß, um starke Korrelationen zwischen den beiden Messzeitpunkten mit einem Fehler erster Art von 5 % und einem Fehler zweiter Art von 20 % nachweisen zu können [101]. Die in Abschnitt 1.9. formulierte Hypothese kann angenommen werden.

#### *4.2.3. Konstruktvalidität des B-STT*

Der Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT-B ist sowohl bei der Gesamtzahl der gesunden Probanden als auch bei den Untersuchungsgruppen der klinischen Studie sehr stark ausgeprägt (konvergente Validität) und stärker als der Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT-A (diskriminante Validität). Dies bleibt in der klinischen Stichprobe auch bei Kontrolle des globalen kognitiven Niveaus bestehen: Ein allgemeines Nachlassen der kognitiven Leistungsfähigkeit hat somit nur einen relativ geringen Anteil an der Assoziation der Testleistungen. Der untersuchte Aspekt der Konstruktvalidität ist erfüllt; der B-STT erfasst damit ähnliche Konstrukte wie der TMT-B. Die in Abschnitt 1.9. formulierte Hypothese kann somit angenommen werden.

Die Höhe der Korrelationen zwischen der Bearbeitungszeit im B-STT und den Parametern der Tests der neuropsychologischen Batterie CERAD-Plus stützt diesen Befund. Die Bearbeitungszeit im B-STT weist nach Korrektur für das globale kognitive Funktionsniveau nur geringe Assoziationen zu den weiteren Testwerten auf. Dies deutet darauf hin, dass sich die zugrunde liegenden Konstrukte vom B-STT und den weiteren Testverfahren der CERAD-Plus unterscheiden.

#### 4.2.4. Kriteriumsvalidität des B-STT

Bezüglich der Bearbeitungszeit konnten signifikante Unterschiede zwischen den Gruppen der klinischen Studie (gesunde Kontrollgruppe, Patienten mit MCI, Patienten mit DAT) nachgewiesen werden, wobei die schnellsten Bearbeitungszeiten von der gesunden Kontrollgruppe erzielt wurden, während die Gruppe der Patienten mit DAT die stärksten Defizite aufwies. Die in Abschnitt 1.9. formulierte Hypothese kann somit angenommen werden.

Die Effektgröße und die Trennschärfe stellen Maße zur Diskriminationsfähigkeit zwischen den Gruppen dar und erleichtern damit als deskriptive Parameter die Interpretation der Gruppenkontraste in Bezug auf die Kriteriumsvalidität und den Vergleich mit dem TMT-B.

Die Effektgrößen der Gruppenzugehörigkeit auf die Bearbeitungszeit im B-STT waren bei allen Kontrasten stark ausgeprägt und in einem jeweils ähnlichen Bereich wie die Effekte auf die Bearbeitungszeit im TMT-B. Die stärksten Effekte finden sich erwartungsgemäß beim Vergleich von KG und DAT. Diese Befunde spiegeln sich auch in den Receiver-Operating-Characteristic-Analysen (ROC-Analysen) wider: Bei allen Gruppenvergleichen zeigen sich gute, dem TMT-B ähnliche Klassifikationsraten. Beim Vergleich von KG und DAT liegt eine nahezu perfekte Klassifikation vor.

#### 4.2.5. Vergleich der Bearbeitungsdauer von B-STT und TMT-B

Der Vergleich der Bearbeitungszeiten in den beiden Testverfahren belegt eine schnellere Bearbeitung des B-STT im Vergleich zum TMT-B. Die zeitlichen Kosten bei der Anwendung des B-STT sind also – für Patient und Untersucher – geringer. Der Effekt des Unterschieds ist überdies sehr stark ausgeprägt. Die in Abschnitt 1.9. formulierte Hypothese kann somit angenommen werden.

Die Ergebnisse des Gruppenvergleichs deuten darüber hinaus auf zwei verschiedene Aspekte hin: Erstens scheint das Ausmaß der Zeitersparnis beim Einsatz des B-STT im Vergleich zum TMT-B im Alter größer zu sein als bei jüngeren Personen. Zweitens scheinen Patienten mit MCI und DAT in stärkerem Ausmaß von der Anwendung des Verfahrens im Sinne einer Verkürzung der Testzeit zu profitieren.

### **4.3. Kritische Betrachtung der Methoden**

In diesem Abschnitt soll erörtert werden, inwieweit die in dieser Untersuchung eingesetzten Methoden geeignet sind, um die Fragestellungen der Studie zu beantworten. Dafür werden kritische Aspekte der Methoden detailliert diskutiert und Implikationen für mögliche Fehler bzw. Limitationen abgeleitet.

#### **4.3.1. Zusammensetzung und Repräsentativität der Stichprobe der gesunden Probanden**

Bedingt durch die Art der Rekrutierung sind die Substichproben der verschiedenen Altersbereiche der gesunden Probanden unterschiedlich umfangreich. Insbesondere für die 31- bis 50-Jährigen stehen im Vergleich zu den anderen Altersbereichen wenige Daten zur Verfügung. Auffällig ist außerdem, dass das allgemeine Bildungsniveau der Probanden über alle Altersgruppen hinweg hoch ausgeprägt war, so dass Personen mit niedrigem Bildungsniveau möglicherweise nicht ausreichend repräsentiert sind. So befanden sich in der gesamten Stichprobe nur achtzehn Personen mit weniger als zwölf Bildungsjahren. Durch den korrelativen und regressionsanalytischen Untersuchungsansatz, bei dem die Assoziationen der Werte über den gesamten Merkmalsbereich untersucht werden, wird dieses Problem jedoch eingeschränkt. Bei der Bewertung der Stichprobengröße und Verteilung muss zudem berücksichtigt werden, dass die Studie keinen Anspruch erhebt, eine Normierung des Verfahrens zu ermöglichen. Bei einer weiterführenden Untersuchung unter dem Gesichtspunkt einer Normierung sollte in jedem Fall auf eine ausgewogenere Repräsentation aller Alters- und Bildungsbereiche geachtet werden.

Für die Repräsentativität der Stichprobe sind weiterhin verschiedene Aspekte der Rekrutierung zu diskutieren:

Personen, die über die Gedächtnissprechstunde des St. Hedwig-Krankenhauses eingeschlossen wurden, hatten sich dort aufgrund von kognitiven Störungen vorgestellt, die ihnen selbst oder anderen Personen aufgefallen waren. Die Bedeutung subjektiver kognitiver Einbußen für die Einschätzung des Risikos, an einer Demenz zu erkranken, ist kontrovers diskutiert worden. In einer metaanalytischen Auswertung zeigte sich ein erhöhtes Risiko für die Entwicklung dementieller Syndrome, teils auch bei unauffälligen kognitiven Leistungen im ersten neuropsychologischen Assessment [107]. Trotz Sicherstellung eines aktuell unauffälligen Status in der psychiatrischen, paraklinischen und neuropsychologischen Diagnostik kann daher nicht ausgeschlossen werden, dass sich ein Teil der Teilnehmenden in einem präklinischen Stadium

---

eines Demenzsyndroms befand und dadurch zum Beispiel in der neuropsychologischen Diagnostik noch formal unauffällige, aber unter einem prämorbidem Funktionsniveau liegende Leistungen erzielte.

Bei den Teilen der Stichprobe, die über Flyer, Aushänge und persönliche Ansprache sowie im Rahmen der Studie zum Arbeitsgedächtnistraining erhoben wurden, muss ein anderer Aspekt beachtet werden: Die Kontrolle der Einschlusskriterien beruhte hierbei auf einer strukturierten Erhebung der Anamnese (siehe Abschnitt 2.1.1.), wobei die zum Ausschluss aus der Studie führenden Kriterien zwar ausführlich erfragt wurden, die Ehrlichkeit und Einsicht der Probanden aber vorausgesetzt werden musste. Bei allen über 50-jährigen Personen wurde das kognitive Funktionsniveau regelmäßig mit dem MMST und in der Arbeitsgedächtnisstudie zusätzlich auch durch eine ausführliche neuropsychologische Untersuchung geprüft. Die Integrität der kognitiven Leistungen ist damit in diesem Teil der Stichprobe besser gesichert. Gleichwohl kann bei den lediglich mit dem MMST untersuchten Personen trotz des konservativ angesetzten Cut-Offs von 27 Punkten das Vorliegen einer leichten kognitiven Störung nicht vollständig ausgeschlossen werden.

Die genannten Schwierigkeiten bei der Generierung der Stichprobe der gesunden Probanden sind prinzipiell auch relevant für ihre Funktion als Kontrollgruppe zu den klinischen Populationen. Hierbei ist jedoch zu beachten, dass von 97 der 107 älteren Probanden eine ausführliche neuropsychologische Charakterisierung mit der Bewertung als kognitiv unauffällig vorlag. Zudem hätte eine Kontamination der Kontrollgruppe mit Patienten mit kognitiven Störungen eher zu einer Verminderung der Gruppenkontraste geführt; beobachtete Gruppenunterschiede spiegeln damit vermutlich genuine Leistungsunterschiede zwischen dem gesunden Alter und den Krankheitsentitäten MCI und DAT wider.

Für den Vergleich der älteren Probanden mit den klinischen Gruppen muss weiterhin berücksichtigt werden, dass trotz gleicher Altersspanne in allen Gruppen das mittlere Alter der Probanden geringfügig unter dem der Patienten lag. Dies ist insofern relevant, als ein Einfluss des Alters auf die Testleistungen nachgewiesen wurde und der Einfluss des Alters bei Verwendung nonparametrischer Analysen bei der Ermittlung der Gruppenunterschiede nicht herausgerechnet werden konnte.

#### 4.3.2. Zusammensetzung und Repräsentativität der klinischen Stichprobe

Alle in die Studie mit der Diagnose DAT eingeschlossenen Patienten erfüllten die Kriterien für eine klinisch wahrscheinliche DAT nach den Kriterien der *National Institute of Neurological and Communicative Disorders and Stroke – The Alzheimer's Disease and Related Disorders Association* (NINCDS-ADRDA; [50]). Die Konkordanz zwischen der Diagnose einer klinisch wahrscheinlichen DAT nach diesen Kriterien und einer sicheren DAT nach neuropathologischer Untersuchung der verstorbenen Patienten wird in der Literatur als hoch angegeben [108]. Dennoch kann bei querschnittlicher Untersuchung im Rahmen der Arbeit nicht ausgeschlossen werden, dass im Einzelfall Patienten mit der Diagnose einer DAT klassifiziert wurden, die bei längsschnittlicher Betrachtung in Frage gestellt werden könnte.

Alle Patienten mit MCI erfüllten die ICD-Kriterien der leichten kognitiven Störung, wobei insbesondere auch das Vorliegen weiterer psychiatrischer Erkrankungen ausgeschlossen wurde. Es ist möglich, dass sich einige der Patienten in einem Übergangsstadium zur beginnenden DAT [62] oder einer anderen Demenzform befanden. Gleichwohl erfolgte die Gruppenzuordnung unter Berücksichtigung des Status zum Zeitpunkt der neuropsychologischen Untersuchung, so dass die Aussagekraft der Ergebnisse diesbezüglich nicht beeinträchtigt ist. Diskutiert werden muss, dass Patienten mit unterschiedlicher neuropsychologischer Phänomenologie (Vergleich Subtypen der MCI, Abschnitt 1.6.3.) für die Untersuchung in eine Gruppe gepoolt wurden. Dementsprechend können keine Aussagen über die Leistungen im B-STT bei verschiedenen Subgruppen der MCI getroffen werden. Dies war allerdings auch nicht Ziel der Untersuchung.

Ein allgemeines Problem bei der Rekrutierung der Patienten bestand darin, dass ausschließlich diejenigen Personen mit dem B-STT untersucht wurden, die in der Lage waren, den TMT-A und TMT-B zu absolvieren. Für die vorliegende Untersuchung ist dies jedoch gerechtfertigt, da der Schwerpunkt auf dem Zusammenhang der Testleistungen sowie dem Vergleich der diagnostischen Validität lag.

#### 4.3.3. Durchführung der Untersuchung

Die Durchführung des B-STT selbst war aufgrund der strengen Befolgung des Anwendungsleitfadens (Abschnitt 2.2.1., Anhang 5.2.) standardisiert. Es lagen jedoch teilweise Unterschiede in der Gesamtdauer der Untersuchung bei den verschiedenen Gruppen vor, die

durch eine unterschiedliche Ausführlichkeit der neuropsychologischen Untersuchung bedingt sind (siehe Abschnitt 2.2., Tabelle 1).

Der B-STT wurde in jedem Fall im Anschluss an den TMT gegen Ende der Untersuchung durchgeführt. Abhängig von der Zahl der zuvor durchgeführten Testverfahren bzw. der Testdauer bis zur Durchführung des B-STT könnten differentielle Ermüdungserscheinungen bei den Versuchspersonen eingetreten sein, die die Leistungen im B-STT akzentuiert haben könnten. Im Rahmen der klinischen Studie, d. h. beim Vergleich der Gruppen KG, MCI und DAT, ist dieser Aspekt jedoch von untergeordneter Bedeutung: Die Mehrheit der in diese Analyse eingeschlossenen Personen durchlief die gesamte neuropsychologische Batterie, wodurch vergleichbare Untersuchungsbedingungen bestanden haben (siehe Abschnitt 3.1.4., Tabelle 7).

#### ***4.4. Der B-STT als Instrument in der neuropsychologischen Diagnostik***

Vor der abschließenden Einschätzung, inwieweit der B-STT als Alternative zum TMT-B eingesetzt werden kann, soll zunächst unter Berücksichtigung der theoretischen Fundierung des Verfahrens, der Befunde der vorliegenden Studie, der diskutierten Limitationen sowie der aktuellen Literatur erörtert werden, ob das Verfahren als Testinstrument für die neuropsychologische Leistungsdiagnostik geeignet ist. Hierzu wird herausgearbeitet, inwieweit der B-STT den Gütekriterien psychodiagnostischer Verfahren genügt.

##### *4.4.1. Hauptgütekriterien*

###### *Objektivität*

Als Objektivität wird die Unabhängigkeit des Untersuchungsergebnisses vom Untersucher bei der Durchführung, Auswertung und Interpretation der Messung bezeichnet [109]. Die Objektivität wurde in der vorliegenden Studie nicht spezifisch untersucht. Unter strenger Manualisierung des Testablaufs (Abschnitt 2.2.1., Anhang 5.2.) können die Durchführungs- und Auswertungsobjektivität vorausgesetzt werden. Eine Interpretation individueller Testleistungen ist gegenwärtig aufgrund noch nicht verfügbarer Normdaten nicht möglich. Nach erfolgter Normierung kann die Interpretationsobjektivität durch Bereitstellung geeigneter Normwerttabellen oder automatisierter Regressionsmodelle sichergestellt werden.

### *Reliabilität*

Unter der Reliabilität eines Verfahrens versteht man seine Messgenauigkeit. Ein Merkmal sollte exakt, also möglichst frei von Messfehlern, erfasst werden [109]. Beim Einsatz von Testverfahren zur Leistungsdiagnostik, und insbesondere bei der Erstellung von Diagnosen mit klinischer Relevanz, ist eine hohe Reliabilität unerlässlich. Die Reliabilität wurde in der vorliegenden Studie als Test-Retest-Reliabilität, also als zeitliche Stabilität, ermittelt. In Bezug auf die Bearbeitungszeiten lag ein starker bis sehr starker Zusammenhang zwischen den Testleistungen vor. Die Reliabilität liegt damit im oberen Bereich der in der Literatur für den TMT-B und den Color Trails Test (CTT) berichteten Werte [87,110–115]. Sie liegt weiterhin innerhalb des Bereichs, der sich metaanalytisch bei neuropsychologischen Verfahren im Allgemeinen gezeigt hat [116]. Für die Bearbeitungszeit im B-STT ist damit bezogen auf die Anforderungen neuropsychologischer Verfahren das Gütekriterium der Reliabilität erfüllt.

### *Validität*

Die Validität eines Verfahrens bezeichnet die Übereinstimmung zwischen dem intendierten und dem tatsächlich erfassten Merkmal [109]. In der vorliegenden Untersuchung wurden die Konstrukt- und die Kriteriumsvalidität geprüft.

Der Aspekt der *Konstruktvalidität* bezieht sich auf die theoretische Fundierung des mit dem Test gemessenen Merkmals [109], hier bestimmt über die Assoziationsstärke zum etablierten TMT. Der stichprobenunabhängig hohe Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT-B weist stark darauf hin, dass die zugrunde liegenden latenten Merkmale übereinstimmend erfasst werden (konvergente Validität). In gleichem Sinne spricht die signifikant geringere Korrelation zwischen B-STT und TMT-A für eine stärkere Unterschiedlichkeit der erfassten Merkmale (diskriminante Validität). Letztere liegt trotz augenscheinlich gemeinsam erforderlicher visuomotorischer Prozesse sowie der bei der Konstruktion des B-STT erreichten gestalterischen Ähnlichkeit zum TMT-A (siehe Abschnitt 2.2.1.) vor.

Dass der Leistung im B-STT ähnliche latente Eigenschaften zugrunde liegen wie der Leistung im TMT-B, wird durch die weiterführenden Analysen gestützt. Bemerkenswert ist insbesondere, dass auch bei Elimination des globalen kognitiven Funktionsniveaus ein hoher Zusammenhang zwischen den Bearbeitungszeiten von B-STT und TMT-B verbleibt; die Relevanz der allgemeinen kognitiven Leistungsfähigkeit scheint also im Vergleich zu den spezifischen

Konstrukten gering zu sein. Auch die geringen Korrelationen zu den verschiedenen Variablen der CERAD-Batterie können als Hinweise auf eine diskriminante Validität dienen; die Aussagekraft ist jedoch aufgrund des augenscheinlich konstruktfernen Charakters der verschiedenen Verfahren (siehe Abschnitt 2.2.2.) eingeschränkt [109].

Die *Kriteriumsvalidität* gibt an, inwieweit das Testergebnis einer Person Rückschlüsse auf ein Außenkriterium, in diesem Fall die Zuordnung zu einer der Diagnosegruppen, zulässt [109]. Für diese Aussage sind die deskriptiven Parameter Effektstärke und Trennschärfe nützlicher als der bloße Nachweis signifikanter Leistungsunterschiede, da hierbei hohe Ausprägungen mit einer höheren Zuordnungsgüte bzw. Diagnosesicherheit einhergehen.

Für die Bearbeitungszeit indizieren die Effektgröße und Trennschärfe eine hohe Diskriminationsfähigkeit des B-STT. Dies betrifft den Vergleich zwischen allen Untersuchungsgruppen (KG, MCI, DAT). Die Bearbeitungszeit im B-STT ist damit bei gleichfalls gegebener Objektivität und Reliabilität ein geeignetes Maß, um in der neuropsychologischen Leistungsdiagnostik zwischen diesen Gruppen zu differenzieren.

#### 4.4.2. Nebengütekriterien

##### *Skalierung*

Skalierung als Gütekriterium beinhaltet, dass die Relation der Leistungen bei verschiedenen Messungen adäquat in den Messwerten wiedergegeben wird. Im Sinne der klassischen Testtheorie ist dabei das Vorliegen mindestens metrisch skalierten Testwerte ein ausreichender Beleg [109]. Die Bearbeitungszeit weist als gemessener Parameter ein metrisches Skalenniveau auf. Zudem wurde durch das Fehlen eines Abbruchkriteriums nach einer bestimmten Bearbeitungsdauer gewährleistet, dass in der Untersuchung das gesamte Merkmalspektrum erfasst werden konnte.

##### *Testökonomie*

Die Wirtschaftlichkeit eines Tests ergibt sich aus dem Verhältnis der verwendeten Ressourcen zum diagnostischen Erkenntnisgewinn [109]. Wie in Abschnitt 4.4.1. dargelegt, ist die Bearbeitungszeit im B-STT ein geeigneter Parameter zur Unterscheidung zwischen gesunden Personen, Patienten mit MCI und Patienten mit DAT. Bezüglich der verwendeten Ressourcen müssen der finanzielle Aufwand und der Zeitaufwand betrachtet werden. Die finanziellen Kosten

entsprechen bei Kopier- bzw. Druckbarkeit des Verfahrens denen des TMT-B. Die zeitlichen Kosten können direkt durch die Bearbeitungszeiten der Tests ermittelt werden. Aus der aktuellen Untersuchung ist ersichtlich, dass der B-STT gruppenübergreifend schneller durchgeführt werden konnte als der TMT-B. Ursache dafür könnte eine durch das Design der Testvorlage bedingte geringere Beanspruchung motorischer und visueller Funktionen sein. Alternativ könnte die schnellere Bearbeitung des B-STT möglicherweise auch durch Reihenfolge- bzw. Übungseffekte bedingt sein: Es wurde nachgewiesen, dass die Bearbeitungszeiten bzw. der Quotient der Bearbeitungszeiten im TMT durch die Reihenfolge der Präsentation der beiden Untertests TMT-A und TMT-B beeinflusst werden [117,118], was auf ein Training der visuellen Suchprozesse zurückgeführt wird [118]. Aufgrund der hohen konzeptuellen Ähnlichkeit zwischen dem B-STT und dem TMT ist es denkbar, dass solche Übungseffekte auch für die Leistungen der Probanden und Patienten im B-STT relevant gewesen sein könnten. Diese könnten sogar ausgeprägter sein als die Effekte innerhalb des TMT, da die Teilnehmer einerseits jeweils zwei „Übungsdurchgänge“ (TMT-A und TMT-B) absolvierten und andererseits mit dem TMT-B stets auch die Aufgabe durchliefen, die als Modell für die Entwicklung des B-STT diente. Die in der Literatur genannten Effekte sind jedoch sehr gering ausgeprägt [117,118], so dass es unwahrscheinlich ist, dass sie für einen wesentlichen Teil des in dieser Studie beobachteten Zeitunterschieds verantwortlich sind.

### *Zumutbarkeit*

Das Kriterium der Zumutbarkeit ist erfüllt, wenn ein Verfahren absolut und in Relation zu seinem Nutzen die Testperson in zeitlicher, psychischer und körperlicher Hinsicht nicht unverhältnismäßig belastet [109]. In der Diagnostik kognitiver Störungen ist die neuropsychologische Leistungsuntersuchung ein unverzichtbarer Baustein [5,50]. Eine Belastung für die Testperson kann sich jedoch durch die Dauer der Untersuchung ergeben. Verfahren, die zu einer Verkürzung der Untersuchung führen, können also Vorteile bezüglich der Zumutbarkeit mit sich bringen.

Wie in den vorherigen Abschnitten dargestellt wurde, hat der B-STT im Vergleich zum TMT-B den Vorteil einer schnelleren Durchführbarkeit bei guter diagnostischer Aussagekraft. Darüber hinaus ergaben sich in der Analyse der Differenzen der Bearbeitungszeiten von B-STT und TMT-B in den verschiedenen Untersuchungsgruppen Hinweise darauf, dass die relative Zeitersparnis mit Zunahme der kognitiven Leistungsminderung ansteigt. Dieser Aspekt ist von besonderer Relevanz, wenn berücksichtigt wird, dass Patienten mit MCI oder DAT regelmäßig

verlängerte Bearbeitungszeiten im TMT-B aufweisen (Vergleich siehe Abschnitte 1.6.2. und 1.6.4.). Die Belastung durch die Untersuchung steigt also bereits durch die Verlängerung der Dauer mit zunehmender Beeinträchtigung der Testperson. Die Zeitersparnis beim B-STT ist dementsprechend ein bedeutender Vorteil gegenüber dem TMT-B und kann zur Entlastung der Patienten in der neuropsychologischen Diagnostik führen.

### *Fairness*

Ein Test erfüllt das Gütekriterium der Fairness, wenn es bei seiner Durchführung oder Interpretation zu keiner systematischen Benachteiligung von Testpersonen aufgrund einer Zugehörigkeit zu bestimmten, z. B. ethnischen oder soziokulturellen, Gruppen kommt [109].

Die Intention bei der Entwicklung des B-STT bestand darin, eine fairere und damit universeller anwendbare Alternative zum TMT-B zu erstellen. Aus theoretischen Erwägungen heraus sowie unter Berücksichtigung der Literatur zu den konzeptuell ähnlichen Verfahren wie CTT und STT [72,91] kann bereits festgestellt werden, dass dieses Ziel erreicht wurde: Der B-STT ist von der Kenntnis des lateinischen Alphabets unabhängig und kann damit, sofern ausreichende Kenntnisse der arabischen Zahlenfolge vorliegen, bei Personen aus Kulturkreisen mit anderen Alphabeten sowie bei Personen mit Analphabetismus eingesetzt werden.

Nicht sicher ist hingegen, ob das Verfahren tatsächlich unabhängig von kulturellen Einflüssen ist. So konnte nachgewiesen werden, dass auch bei gänzlich nonverbalen neuropsychologischen Testverfahren interkulturelle Performanzunterschiede auftreten [119]. In ähnlicher Weise zeigten sich sowohl im TMT als auch im Color Trails Test Leistungsunterschiede zwischen verschiedenen kulturellen Gruppen [41,82,83]. Erklärungsansätze für Kulturunterschiede bei neuropsychologischen Testverfahren fokussieren auf die Qualität der Ausbildung in verschiedenen Kulturen [120], kulturspezifische Einstellungen sowie eine kulturabhängige Familiarität mit Leistungstests [82]. Basierend auf diesen Befunden muss vermutet werden, dass unabhängig von der Anwendbarkeit des B-STT auf die Erhebung kulturspezifischer Normdaten nicht verzichtet werden kann.

### *Normierung*

Als Normierung bezeichnet man die Erstellung eines Bezugssystems, welches die Einordnung und Interpretation individueller Testwerte gestattet. Dies wird durch Sammlung einer ausreichend großen Datenmenge an einer Eichstichprobe unter Berücksichtigung einer

ausreichenden Repräsentation relevanter demografischer Merkmale, wie z. B. Alter und Bildung, bewerkstelligt [109]. Als ausreichend erachtet wird dabei eine Zellengröße von mindestens 50–75 Personen [121].

In der aktuellen Studie wurden, da es sich um eine vorläufige Validierung handelte, keine Normdaten generiert. Die bei den gesunden Probanden gewonnenen Daten sollten bis zum experimentellen Ausschluss relevanter Übungseffekte nicht in eine spätere Normierung einbezogen werden, da für die Praxis ein anderer Untersuchungsablauf anzustreben ist, in dem der B-STT anstelle des TMT-B verwendet wird. Dennoch liefern die vorliegenden Daten wertvolle Hinweise für die Planung einer späteren Normierung. Insbesondere lässt sich aus den Untersuchungen zu den Einflüssen demografischer Parameter auf die Testleistungen ableiten, dass die einzelnen Alterszellen möglichst eng gehalten werden sollten, während die Einflüsse geringer Bildungsunterschiede weniger relevant und separate geschlechtsspezifische Normen auf Basis der aktuellen Daten nicht erforderlich sind. Die Anwendung eines Regressionsmodells könnte dabei, analog zur Baseler Norm des TMT [11], zu einer genaueren Darstellung der Einflüsse führen. Wie zuvor dargelegt, sollte zudem geprüft werden, ob kulturspezifische Normwerte erforderlich sind.

#### ***4.5. Der B-STT als Alternative zum TMT-B***

Wie in Abschnitt 4.4. diskutiert, ist der B-STT ein Verfahren, das den geforderten Gütekriterien psychologischer Testverfahren genügt bzw. nach Durchführung einer Normierung voraussichtlich genügen wird. In diesem Abschnitt soll diskutiert werden, ob er darüber hinaus eine geeignete Alternative zum TMT-B darstellt.

Ein Beleg für die Analogie der beiden Verfahren sind die sowohl bei gesunden Personen als auch in der klinischen Population beobachteten hohen Korrelationen der Bearbeitungszeiten. Diese erreichen nahezu das Ausmaß der in der Literatur angegebenen Assoziationen zwischen dem TMT-B und seinen drei durch Spiegelung bzw. Rotation der Testvorlage erzeugten Parallelförmigen [122].

Ein weiterer Beleg ist die vergleichbare diagnostische Klassifikationsleistung zwischen den Gruppen KG, MCI und DAT. Hierbei darf nicht vergessen werden, dass bei allen in der Gedächtnisprechstunde untersuchten Personen die Leistung im TMT-B selbst als Kriterium für

die Einordnung in die Diagnosegruppen berücksichtigt wurde, so dass eine hohe Kriteriumsvalidität die Folge einer hohen Konstruktvalidität darstellen kann.

Die Befunde der aktuellen Studie zur Testökonomie und Zumutbarkeit verweisen zudem darauf, dass der B-STT dem TMT-B auch außerhalb des primär intendierten Einsatzgebietes, der Untersuchung von Patienten ohne Kenntnis des lateinischen Alphabets, nicht nur gleichwertig, sondern überlegen sein könnte.

#### ***4.6. Der B-STT im Vergleich zu den bereits veröffentlichten Alternativen zum TMT-B***

Nach Herausstellung der Eignung des B-STT als vorteilhafte Alternative zum TMT-B soll an dieser Stelle noch diskutiert werden, inwiefern Vor- oder möglicherweise Nachteile gegenüber den zuvor veröffentlichten und als Alternativen zum TMT diskutierten Verfahren Color Trails Test (CTT), Shape Trail Test (STT) und Trail Making Test Black & White (TMT-B&W) bestehen. Beachtet werden muss, dass ein Vergleich spezifischer Testkennwerte nur orientierend erfolgen kann, da sich die Untersuchungsbedingungen und die Zusammensetzung der Stichproben in den zitierten Studien teils deutlich von der vorliegenden Arbeit unterscheiden.

##### *4.6.1. Color Trails Test*

Der wichtigste Vorteil des CTT im Vergleich zum B-STT liegt in der langjährigen Anwendung des Verfahrens und dem Vorliegen von Normwerten [23,72]. Außerdem ist er unter den Alternativen das einzige Verfahren, für das eine Angabe der Test-Retest-Reliabilität vorliegt. Die Schwankungsbreite spricht dabei jedoch teilweise nur für eine befriedigende zeitliche Stabilität [72,87].

Bezüglich der Validität liegen offensichtliche Einschränkungen vor. Zum einen sind die Assoziationen zwischen dem TMT-B und dem CTT-2 überwiegend unbefriedigend [81,87,89,90]. Zum anderen scheint die nicht ausreichende Klassifikationsleistung des CTT-2 beim Vergleich zwischen gesunden Probanden und Patienten mit leichter kognitiver Störung [86] auf einen verglichen mit dem B-STT geringeren diagnostischen Erkenntnisgewinn zu verweisen.

Weiterhin indizieren die in einer aktuellen Untersuchung durchweg höheren Bearbeitungszeiten im CTT-2 [86] im Vergleich zum B-STT (siehe Tabelle 12) Nachteile bezüglich der Testökonomie und Zumutbarkeit, insbesondere bei auch unzureichender Klassifikationsleistung. Die im Vergleich längere Bearbeitungszeit könnte durch das Vorliegen von Distraktoren oder

Eigenschaften der Itemplatzierung bedingt sein. Ein letzter Nachteil im Vergleich mit dem B-STT besteht in der Farbigkeit des Materials, wodurch sich höhere Kosten beim Erstellen der Testvorlagen ergeben.

#### 4.6.2. *Shape Trail Test*

Ein Vorteil des STT gegenüber dem B-STT liegt in der durch die große Stichprobe bedingten exzellenten klinischen Validierung bezüglich der Differenzierung zwischen gesunden Personen und Patienten mit leichter kognitiver Störung oder Demenz bei Alzheimerkrankheit [91]. Dies wird belegt durch ROC-Analysen beim Vergleich von gesunden Personen mit Patienten mit DAT, wobei die Klassifikationsgüte mit einer *area under the curve* von 0,82–0,91 unter den in der aktuellen Studie ermittelten Werten für den B-STT liegt. ROC-Analysen für Patienten mit MCI werden nicht präsentiert.

Nachteilig ist, dass sich die Charakterisierung der demografischen Eigenschaften, zumindest in der zugänglichen Literatur, auf den Altersbereich von 50–90 Jahren beschränkt. Weitere Einschränkungen bestehen hinsichtlich der Konstruktvalidität. Eine Untersuchung des Zusammenhangs zwischen dem STT-B und dem TMT-B erfolgte weder in den chinesischen Veröffentlichungen noch in der in englischer Sprache veröffentlichten Validierungsstudie [91–93]. Zhao et al. (2013; [91]) geben Korrelationen zum CTT an, die jedoch wenig Aufschluss über die Analogie bzw. Konstruktvalidität zum TMT-B erlauben.

Hinsichtlich der Testökonomie und Zumutbarkeit muss außerdem wie beim CTT-2 festgestellt werden, dass die mittleren Bearbeitungszeiten im STT-B augenscheinlich deutlich über denen des B-STT liegen (siehe Tabelle 12).

#### 4.6.3. *Trail Making Test Black & White*

Für den Vergleich zum TMT-B&W liegt die bisher einzige Veröffentlichung zu diesem Testverfahren vor, die eine Untersuchung der Konstrukt- und Kriteriumsvalidität beinhaltet [94]. Bezüglich der Konstruktvalidität verweist die starke Korrelation auf einen hohen Zusammenhang zwischen TMT-B und TMT-B&W-B. Der Anteil der Varianzaufklärung indiziert jedoch auf deskriptiver Ebene einen geringeren Zusammenhang als die in dieser Studie beobachtete Assoziation zwischen TMT-B und B-STT. Bei der Angabe der Leistungsunterschiede zwischen den Untersuchungsgruppen (Kontrollgruppe, Patienten mit leichter kognitiver Störung, Patienten mit Demenz bei Alzheimerkrankheit) wurden signifikante

Unterschiede zwischen den Gruppen berichtet. Eine genauere Quantifizierung anhand von Effektstärken oder Trennschärfen wurde nicht vorgenommen.

Wie schon beim CTT-2 und STT-B fallen auch beim TMT-B&W-B gruppenübergreifend lange Bearbeitungszeiten auf (siehe Tabelle 12); auch hier scheint der B-STT einen Vorteil bezüglich der Testökonomie und Zumutbarkeit zu haben.

**Tabelle 12: Gegenüberstellung der Befunde zu den Bearbeitungszeiten im CTT-2, im STT-B, im TMT-B&W-B sowie im B-STT bei gesunden Personen (KG), Patienten mit MCI und Patienten mit DAT**

Studie		KG	MCI	DAT
CTT-2 [85]	Personen	28		75
	Alter M (SD)	70,2 (7,9)		77,7 (6,9)
	Bildung M (SD)	13,6 (3,4)		11,3 (2,4)
	CTT-2 M (SD)	107,3 (35,1)		220,9 (69,9)
CTT-2 [86]	Personen	15	8	9
	Alter M (SD)	71,1 (6,9)	73,8 (8,7)	74,7 (8,2)
	Bildung M (SD)	12,0 (3,6)	10,4 (4,3)	8,3 (2,9)
	CTT-2 M (SD)	110,0*	129,5*	210,0*
STT-B [91]**	Personen	301	264	125
	Alter M (SD)	73,5 (5,5)	73,9 (4,6)	74,6 (4,1)
	Bildung M (SD)	14,2 (1,8)	14,4 (1,8)	14,0 (1,6)
	STT-B M (SD)	157,5 (49,7)	219,2 (91,9)	285,0 (123,2)
TMT-B&W-B [94]***	Personen	28	42	19
	Alter M (SD)	62,6 (6,2)	69,1 (7,5)	74,1 (8,1)
	Bildung M (SD)	11,61 (6,2)	8,5 (6,6)	8,5 (6,6)
	TMT-B&W-B M (SD)	126,7 (9,1)	154,5 (7,1)	209,2 (10,6)
B-STT	Personen	107	82	36
	Alter M (SD)	71,3 (4,1)	73,6 (4,1)	73,9 (3,7)
	Bildung M (SD)	14,6 (2,6)	13,9 (2,7)	13,8 (2,6)
	B-STT M (SD)	58,8 (18,3)	92,8 (43,2)	149,2 (70,8)

Angabe der Bearbeitungszeiten der Tests in Sekunden; Angabe von Alter und Bildungsdauer in Jahren

\* Keine Angaben von Standardabweichungen in der Publikation

\*\* Daten für Untersuchungsgruppe 4 (Alter > 65 Jahre, Ausbildung > 12 Jahre)

\*\*\* Demografische Daten der gesamten Stichprobe; der Test wurde jedoch nur von einer Subgruppe absolviert

#### **4.7. Zusammenfassung und Ausblick**

In der vorliegenden Arbeit erfolgte eine erste Charakterisierung und Validierung des Berliner Shape Trail Tests (B-STT), welcher als Alternative zum etablierten Trail Making Test B (TMT-B) konzipiert wurde. Für die Untersuchung wurde der B-STT an 235 gesunden Probanden verschiedener Altersgruppen sowie an 82 Patienten mit leichter kognitiver Störung und 36 Patienten mit Demenz bei Alzheimerkrankheit eingesetzt. Als zentrale Befunde zeigten sich im B-STT bezüglich des Leistungsparameters Bearbeitungszeit eine gute Test-Retest-Reliabilität, stichprobenübergreifend hohe Korrelationen zum TMT-B, eine gute diagnostische Klassifikationsleistung zwischen gesunden Probanden, Patienten mit leichter kognitiver Störung und Patienten mit Demenz bei Alzheimerkrankheit sowie ein zeitlicher Vorteil bei der Bearbeitung im Vergleich zum TMT-B. Die Studie konnte damit deutlich belegen, dass der B-STT ein geeignetes Instrument in der neuropsychologischen Leistungsdiagnostik und darüber hinaus eine im Allgemeinen gleichwertige und zumindest unter testökonomischen Aspekten sogar überlegene Alternative zum TMT-B darstellt. Im Vergleich zu den bisher verfügbaren Alternativen ist der B-STT gut charakterisiert und weist Vorteile insbesondere in Bezug auf die Konstruktvalidität, die Testökonomie und die Zumutbarkeit auf. Die Ergebnisse der vorliegenden Arbeit bereiten den späteren Einsatz des B-STT als eine schriftfreie Alternative zum TMT-B vor, die unabhängiger vom sprachlichen Hintergrund sowie bei Vorliegen eines Analphabetismus eingesetzt werden kann.

Der B-STT wurde mit dem Ziel konzipiert, eine Alternative zum TMT-B für die neuropsychologische Untersuchung von Personen mit Analphabetismus und Personen ohne ausreichende Kenntnisse des lateinischen Alphabets darzustellen. Die aktuelle Arbeit stellt mit dem Beleg der allgemeinen Testgütekriterien und dem Nachweis einer hohen Übereinstimmung der erfassten Konstrukte von B-STT und TMT-B bei einer deutschsprachigen Stichprobe nur die Vorarbeit zur Untersuchung dieses Ziels dar. Ausstehend sind weiterführende Untersuchungen, die die Übereinstimmung von B-STT und TMT-B und die Eignung des B-STT bei der eigentlichen Zielpopulation, z. B. im Einsatz bei Personen aus verschiedenen Kulturkreisen, belegen. Hierfür könnte der B-STT analog zur vorliegenden Arbeit gemeinsam mit einem dem jeweiligen Alphabet angepassten TMT-B bei gesunden Personen und Patienten mit kognitiven Störungen angewandt werden.

Bezüglich der Eignung des B-STT bei der Differenzierung kognitiver Störungen liefert die aktuelle Untersuchung nur Belege für den Vergleich zwischen gesundem Altern und kognitiven

---

Einbußen bei Vorliegen einer leichten kognitiven Störung bzw. einer Demenz bei Alzheimerkrankheit. Diese Datenbasis sollte ergänzt werden um eine Charakterisierung anderer Krankheitsentitäten mit dem B-STT. Dabei sollte sowohl an andere Demenzformen, wie z. B. vaskuläre Demenz unterschiedlicher Ätiologie und die verschiedenen Formen der frontotemporalen Demenz, als auch an psychiatrische Erkrankungen wie Major Depression oder Schizophrenie gedacht werden.

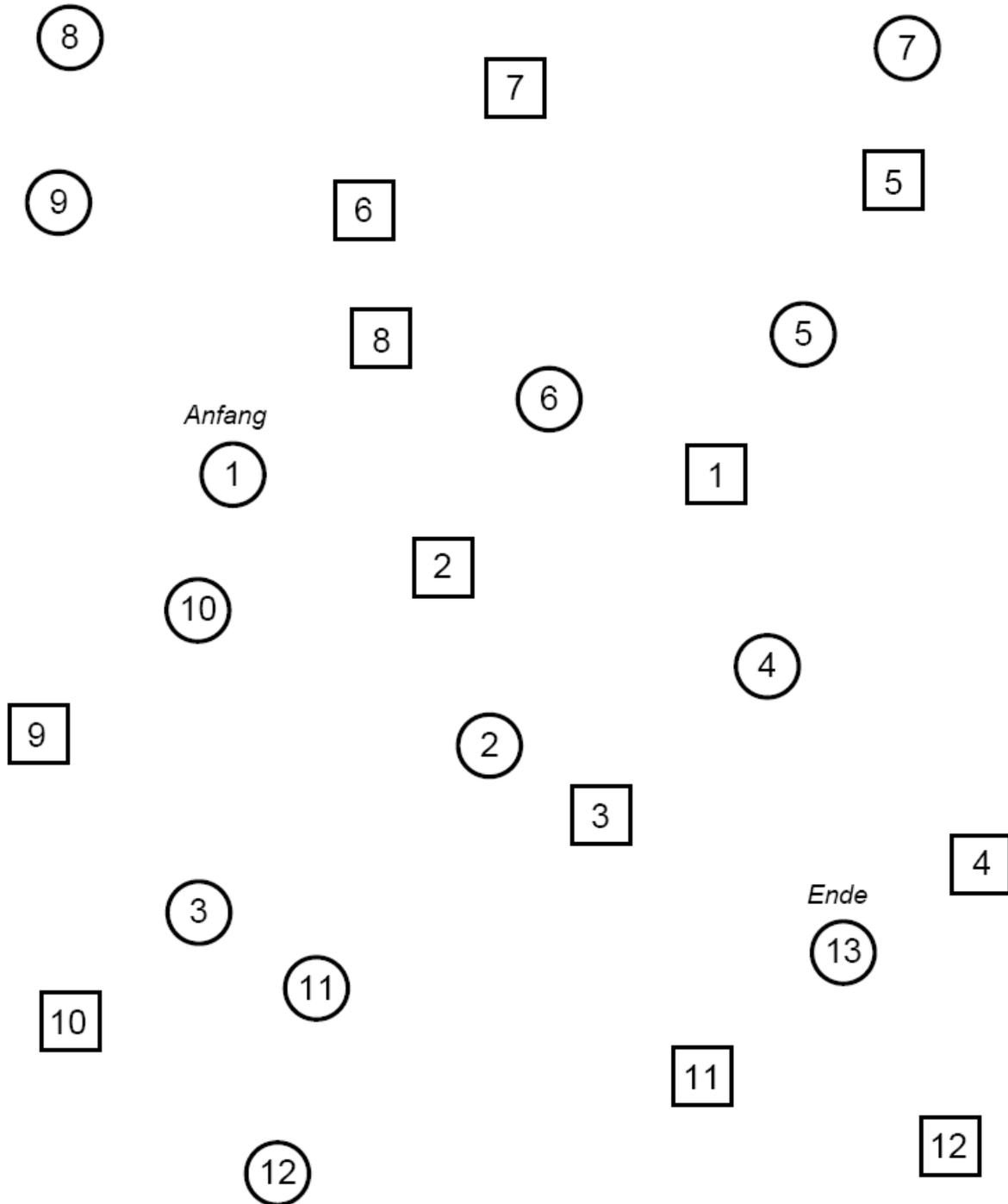
Vor einer systematischen Anwendung bei der klinisch-neuropsychologischen Diagnostik kognitiver Störungen ist eine formelle Normierung des Verfahrens mithilfe einer ausreichend großen Eichstichprobe (siehe Abschnitt 4.4.2.) essentiell. Da der B-STT als Alternative zum TMT-B vorgesehen ist, sollte er dabei anstelle dessen nach dem TMT-A angewendet werden. Bei der Erstellung der Normwerte sollten Alter, Bildungsniveau und natürlich der kulturelle Hintergrund der Testperson berücksichtigt werden.

In der aktuellen Studie wurde aus der hohen Konstruktvalidität geschlossen, dass der B-STT latente exekutive Funktionen in ähnlicher Weise wie der TMT-B erfasst. Eine ergänzende Charakterisierung des B-STT sollte faktorenanalytisch durchgeführt werden. Eine der Faktorenanalyse zugrunde liegende Testbatterie sollte dabei Verfahren zur Erfassung verschiedener exekutiver Funktionen, aber auch Aufmerksamkeitsfunktionen sowie visueller und motorischer Komponenten beinhalten.

Mit dem B-STT liegt ein Verfahren vor, welches einen vielversprechenden Ansatz für eine schriftunabhängige neuropsychologische Untersuchung darstellt und das sich darüber hinaus als besonders ökonomisch erwiesen hat. Die weitere Charakterisierung und eine Normierung des Verfahrens können zur Vereinheitlichung und Verbesserung der Untersuchung von Patienten mit kognitiven Störungen beitragen.

## 5. Anhang

### 5.1. Arbeitsblatt Berliner Shape Trail Test



## 5.2. Testinstruktionen

### 5.2.1. Trail Making Test A (TMT-A)

Übungsblatt: (*Das Blatt wird vorgelegt.*) „Bei dieser Aufgabe sehen Sie hier Kreise mit Zahlen von 1 bis 8. Ich möchte, dass Sie die Kreise so schnell wie möglich in der richtigen Reihenfolge miteinander verbinden (*demonstrieren, ohne tatsächlich zu zeichnen*). Bitte machen Sie dabei gerade Striche und setzen Sie den Stift möglichst nicht ab. Bitte starten Sie.“

Arbeitsblatt: (*Das Blatt wird um 180° rotiert vorgelegt.*) „Das war nur eine Übung. Jetzt kommt die eigentliche Aufgabe. Das Prinzip ist das gleiche; ich werde jetzt aber die Zeit stoppen. Sie sollen also so schnell wie möglich vorgehen und den Stift möglichst nicht absetzen. Ich zeige Ihnen gleich den Anfang, und dann geht es sofort los. (*Das Blatt wird umgedreht, und der Anfang wird gezeigt*). Hier ist der Anfang, los geht's ...“

### 5.2.2. Trail Making Test B (TMT-B)

Übungsblatt: (*Das Blatt wird vorgelegt.*) „Bei dieser Aufgabe sehen Sie in den Kreisen Zahlen und Buchstaben. Ich möchte, dass Sie die Kreise so schnell wie möglich miteinander verbinden, so dass sich Zahlen und Buchstaben abwechseln und die richtige Reihenfolge eingehalten wird. Also 1–A–2–B und so weiter (*demonstrieren, ohne tatsächlich zu zeichnen*). Setzen Sie den Stift möglichst nicht ab. Bitte beginnen Sie.“

Arbeitsblatt: (*Das Blatt wird um 180° rotiert vorgelegt.*) „Das war die Übung. Jetzt kommt die eigentliche Aufgabe. Das Prinzip ist das gleiche: Zahlen und Buchstaben sollen sich abwechseln. Arbeiten Sie so schnell Sie können und setzen Sie den Stift möglichst nicht ab. Ich zeige Ihnen gleich den Anfang, und dann geht es sofort los. (*Das Blatt wird umgedreht, und der Anfang wird gezeigt*). Hier ist der Anfang, los geht's ...“

### 5.2.3. Berliner Shape Trail Test (B-STT)

Übungsblatt: (*Das Blatt wird vorgelegt.*) „Bei dieser Aufgabe sehen Sie Zahlen in Kreisen und Vierecken. Jede Zahl kommt dabei zweimal vor (*zeigen*). Bitte verbinden Sie die Formen so, dass jede Zahl zweimal hintereinander kommt und sich Kreis und Viereck dabei abwechseln, also Kreis-1–Viereck-1–Kreis-2–Viereck-2 und so weiter (*demonstrieren, ohne tatsächlich zu zeichnen*). Setzen Sie den Stift dabei möglichst nicht ab. Bitte starten Sie.“

Arbeitsblatt: (*Das Blatt wird um 180° rotiert vorgelegt.*) „Das war die Übung. Jetzt kommt die eigentliche Aufgabe. Das Prinzip ist das gleiche. Denken Sie daran, dass jede Zahl zweimal vorkommt und sich Kreis und Viereck abwechseln. Arbeiten Sie so schnell Sie können und setzen Sie den Stift möglichst nicht ab. Ich zeige Ihnen gleich den Anfang, und dann geht es sofort los. (*Das Blatt wird umgedreht, und der Anfang wird gezeigt.*) Hier ist der Anfang, los geht's ...“

#### 5.2.4. Instruktionen beim Auftreten von Fehlern

Als Fehler bezeichnet wird jede vollendete falsche Verbindung. Die Versuchsperson wird sofort auf den Fehler hingewiesen: „Achtung! Denken Sie daran, dass ...“ Die weitere Formulierung hängt ab vom Test sowie von der Art des Fehlers (siehe Abschnitt 1.2.; Definition der Fehlertypen für den B-STT analog zum TMT-B):

- Sequenz- und Proximitätsfehler: „... Sie die richtige Reihenfolge einhalten müssen.“
- TMT-B, Perseverationsfehler: „... sich Zahlen und Buchstaben abwechseln müssen.“
- B-STT, Perseverationsfehler: „... sich Kreis und Viereck abwechseln müssen.“

Die Fehlerinstruktion wird beendet mit „Gehen Sie hierher (*Das letzte richtige Zeichen wird gezeigt.*) und zeichnen Sie zum nächsten richtigen Zeichen.“

Ist der Patient nicht in der Lage, die Aufgabe mit der Fehlerinstruktion fortzusetzen, wird eine der folgenden Zusatzinstruktionen dargeboten:

- TMT-A: “Sie müssen die Zahlen in aufsteigender Reihenfolge verbinden.”
- TMT-B: “Zahlen und Buchstaben müssen sich abwechseln, und die richtige Reihenfolge muss eingehalten werden.”
- B-STT: “Jede Zahl kommt zweimal nacheinander vor. Kreis und Viereck wechseln sich dabei ab.”

### 5.3. Altersunterschiede innerhalb der gesunden Probanden

**Tabelle 13: Testgrößen für den Altersvergleich der sechs Subgruppen der gesunden Probanden**

Altersgruppe	t-Test
21–30 Jahre vs. 31–40 Jahre	t(46)=−13,118, p=.000
31–40 Jahre vs. 41–50 Jahre	t(21)=−8,371, p=.000
41–50 Jahre vs. 51–60 Jahre	t(43)=−10,776, p=.000
51–60 Jahre vs. 61–70 Jahre	t(114)=−14,082, p=.000
61–70 Jahre vs. 71–80 Jahre	t(140)=−19,227, p=.000

Alle p-Werte sind lokal mit der Bonferroni-Holm-Prozedur adjustiert.

### 5.4. Neuropsychologische Charakteristika in der klinischen Studie – Prüfung der Voraussetzungen für die ANOVA

**Tabelle 14: Prüfung der Varianzhomogenität und Homoskedastizität der Residuen für die neuropsychologischen Parameter innerhalb der klinischen Studie (Vergleich KG, MCI, DAT)**

Test	Kolmogorov-Smirnov-Test	Levene-Test
MMST	z=2,492, p=.000	F(2,222)=11,548, p=.000
WL <sub>Ges</sub> *	z=1,062, p=.210	F(2,212)=0,945, p=.390
WL <sub>Abr</sub> *	z=1,276, p=.077	F(2,212)=2,969, p=.054
WL <sub>Rek</sub>	z=3,732, p=.000	F(2,212)=60,857, p=.000
BNT	z=4,226, p=.000	F(2,212)=12,526, p=.000
VKon**	z=2,959, p=.000	F(2,212)=1,359, p=.259
Log(VKon)**	z=3,507, p=.000	F(2,212)=1,428, p=.242
FigG	z=1,442, p=.031	F(2,212)=19,794, p=.000
Flu-T	z=1,236, p=.094	F(2,212)=3,356, p=.037
Flu-S*	z=0,733, p=.656	F(2,210)=0,467, p=.628
TMT-A Zeit	z=1,409, p=.038	F(2,222)=20,486, p=.000
TMT-B Zeit	z=1,981, p=.001	F(2,222)=43,080, p=.000

\* Voraussetzungen der ANOVA erfüllt

\*\* Herstellung einer Normalverteilung auch durch Log<sub>10</sub>-Transformation nicht möglich

### 5.5. Neuropsychologische Charakteristika in der klinischen Studie – Gruppenunterschiede

Tabelle 15: Prüfgrößen und Signifikanzwerte für die Vergleiche der neuropsychologischen Charakteristika der Untersuchungsgruppen (KG, MCI, DAT)

Verfahren	Globaltest: Prüfgrößen, Signifikanzwerte	Lokaltest: Prüfgrößen, Signifikanzwerte
MMST*	$\chi^2(2)=70,730$ , p=.000	$Z_{KG.MCI}=-4,143$ , p=.000; $Z_{KG.DAT}=-7,754$ , p=.000; $Z_{MCI.DAT}=-5,933$ , p=.000
WL <sub>Ges</sub> **	F(2)=76,253, p=.000	t(177) <sub>KG.MCI</sub> =3,947, p=.000; t(131) <sub>KG.DAT</sub> =13,141, p=.000; t(116) <sub>MCI.DAT</sub> =8,604, p=.000
WL <sub>Abr</sub> **	F(2)=106,022, p=.000	t(145,429) <sub>KG.MCI</sub> =-6,395, p=.000; t(131) <sub>KG.DAT</sub> =16,173, p=.000; t(116) <sub>MCI.DAT</sub> =8,415, p=.000
WL <sub>Rek</sub> *	$\chi^2(2)=64,330$ , p=.000	$Z_{KG.MCI}=-2,991$ , p=.003; $Z_{KG.DAT}=-7,712$ , p=.000; $Z_{MCI.DAT}=-5,750$ , p=.000
BNT*	$\chi^2(2)=20,326$ , p=.000	$Z_{KG.MCI}=-2,768$ , p=.006; $Z_{KG.DAT}=-4,409$ , p=.000; $Z_{MCI.DAT}=-2,180$ , p=.029
VKon*	$\chi^2(2)=2,659$ , p=.265	nicht durchgeführt
FigG*	$\chi^2(2)=73,680$ , p=.000	$Z_{KG.MCI}=-5,040$ , p=.000; $Z_{KG.DAT}=-7,746$ , p=.000; $Z_{MCI.DAT}=-5,449$ , p=.000
Flu-T*	$\chi^2(2)=53,685$ , p=.000	$Z_{KG.MCI}=-4,182$ , p=.000; $Z_{KG.DAT}=-7,068$ , p=.000; $Z_{MCI.DAT}=-3,846$ , p=.000
Flu-S**	F(2)=24,739, p=.000	t(176) <sub>KG.MCI</sub> =5,597, p=.000; t(130) <sub>KG.DAT</sub> =5,820, p=.000; t(114) <sub>MCI.DAT</sub> =1,646, p=.102
TMT-A Zeit*	$\chi^2(2)=46,862$ , p=.000	$Z_{KG.MCI}=-5,331$ , p=.000; $Z_{KG.DAT}=-5,709$ , p=.000; $Z_{MCI.DAT}=-2,075$ , p=.038
TMT-B Zeit*	$\chi^2(2)=92,414$ , p=.000	$Z_{KG.MCI}=-7,265$ , p=.000; $Z_{KG.DAT}=-8,073$ , p=.000; $Z_{MCI.DAT}=-3,609$ , p=.000

\* Globaltest: Kruskal-Wallis-Test; Lokaltest: Wilcoxon-Mann-Whitney-Test

\*\* Globaltest: ANOVA; Lokaltest: t-Test für unabhängige Stichproben

---

## Literaturverzeichnis

1. Niebuhr A, Stiller S. Demographischer Wandel und Arbeitsmärkte in Norddeutschland. *Wirtschaftsdienst* 2005;85:326–332.
2. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, Statistisches Bundesamt (Hrsg.). Herausforderungen des demografischen Wandels. Expertise im Auftrag der Bundesregierung. Paderborn: Bonifatius GmbH Buch-Druck-Verlag, 2011.
3. Saß A-C, Wurm S, Ziese T. Somatische und psychische Gesundheit. In: *Gesundheit und Krankheit im Alter. Eine gemeinsame Veröffentlichung des Statistischen Bundesamtes, des Deutschen Zentrums für Altersfragen und des Robert Koch-Instituts*. Berlin: Robert Koch-Institut, 2009.
4. Weyerer S. Altersdemenz. Gesundheitsberichtserstattung des Bundes. Heft 28. Berlin: Robert Koch-Institut, 2005.
5. Eschweiler GW, Leyhe T, Klöpffer S, Hüll M. Neue Entwicklungen in der Demenzdiagnostik. *Dtsch Arztebl* 2010;107:677–683.
6. Tischler L, Petermann F. Trail Making Test (TMT). *Z Psychiatr Psychol Psychother* 2010;58:79–81.
7. Soukup VM, Ingram F, Grady JJ, Schiess MC. Trail Making Test: Issues in normative data selection. *Appl Neuropsychol* 1998;5:65–73.
8. Fine EM, Delis DC, Holdnack J. Normative adjustments to the D-KEFS Trail Making Test: Corrections for education and vocabulary level. *Clin Neuropsychol* 2011;25:1331–1344.
9. Weissenborn K, Rückert N, Hecker H, Manns MP. The Number Connection Tests A and B: interindividual variability and use for the assessment of early hepatic encephalopathy. *J Hepatol* 1998;28:646–653.
10. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. *Nat Protoc* 2006;1:2277–2281.
11. Memory Clinic Basel (Hrsg.). CERAD-Plus: The Consortium to Establish a Registry for Alzheimer's Disease. Basel: 2005.
12. Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E, Crane PK, Gibbons LE. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): An expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav* 2012;6:489–501.
13. Drane DL, Yuspeh RL, Huthwaite JS, Klingler LK. Demographic characteristics and normative observations for derived-Trail Making Test indices. *Neuropsychiatry Neuropsychol Behav Neurol* 2002;15:39–43.

14. Stuss DT, Bisschop SM, Alexander MP, Levine B, Katz D, Izukawa D. The Trail Making Test: A study in focal lesion patients. *Psychol Assess* 2001;13:230–239.
15. Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Arch Clin Neuropsychol* 2008;23:129–137.
16. Sanchez-Cubillo I, Perianez JA, Adrover-Roig D, Rodriguez-Sanchez JM, Rios-Lago M, Tirapu J, Barcelo F. Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J Int Neuropsychol Soc* 2009;15:438–450.
17. Groff MG, Hubble LM. A factor analytic investigation of the Trail Making Test. *Clin Neuropsychol* 1981;3:11–13.
18. Rasmusson XD, Zonderman AB, Kawas C, Resnick SM. Effects of age and dementia on the Trail Making Test. *Clin Neuropsychol* 1998;12:169–178.
19. Ricker JH, Axelrod BN. Analysis of an oral paradigm for the Trail Making Test. *Assessment* 1994;1:47–52.
20. Hirota C, Watanabe M, Tanimoto Y, Kono R, Higuchi Y, Kono K. A cross-sectional study on the relationship between the Trail Making Test and mobility-related functions in community-dwelling elderly. *Nihon Ronen Igakkai Zasshi* 2008;45:647–654.
21. Corrigan JD, Hinkeldey NS. Relationships between parts A and B of the Trail Making Test. *J Clin Psychol* 1987;43:402–409.
22. Ríos M, Periañez JA, Muñoz-Céspedes JM. Attentional control and slowness of information processing after severe traumatic brain injury. *Brain Inj* 2004;18:257–272.
23. Mitrushina MN, Boone KB, Razani J, D'Elia LF. *Handbook of normative data for neuropsychological assessment* (2nd edition). New York: Oxford University Press, 2005.
24. Rodewald K, Bartolovic M, Debelak R, Aschenbrenner S, Weisbrod M, Roesch-Ely D. Eine Normierungsstudie eines modifizierten Trail Making Tests im deutschsprachigen Raum. *Z Neuropsychol* 2012;23:37–48.
25. Crowe SF. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. *J Clin Psychol* 1998;54:585–591.
26. Langenecker SA, Zubieta J-K, Young EA, Akil H, Nielson KA. A task to manipulate attentional load, set-shifting, and inhibitory control: convergent validity and test-retest reliability of the Parametric Go/No-Go Test. *J Clin Exp Neuropsychol* 2007;29:842–853.
27. Arbuthnott K, Frank J. Trail making test, part B as a measure of executive control: Validation using a set-switching paradigm. *J Clin Exp Neuropsychol* 2000;22:518–528.
28. Stanczak DE, Lynch MD, McNeil CK, Brown B. The Expanded Trail Making Test: Rationale, development, and psychometric properties. *Arch Clin Neuropsychol* 1998;13:473–487.

29. Gaudino EA, Geisler MW, Squires NK. Construct validity in the Trail Making Test: What makes part B harder? *J Clin Exp Neuropsychol* 1995;17:529–535.
30. Fossum B, Holmberg H, Reinvang I. Spatial and symbolic factors in performance on the Trail Making Test. *Neuropsychology* 1992;6:71–75.
31. Hamdan AC, Hamdan EMLR. Effects of age and education level on the Trail Making Test in a healthy Brazilian sample. *Psychol Neurosci* 2009;2:199–203.
32. Bezdicek O, Motak L, Axelrod BN, Preiss M, Nikolai T, Vyhnalek M, Poreh A, Ruzicka E. Czech version of the Trail Making Test: Normative data and clinical utility. *Arch Clin Neuropsychol* 2012;27:906–914.
33. Periañez JA, Ríos-Lago M, Rodríguez-Sánchez JM, Adrover-Roig D, Sánchez-Cubillo I, Crespo-Facorro B, Quemada JI, Barceló F. Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Arch Clin Neuropsychol* 2007;22:433–447.
34. Zalonis I, Kararizou E, Triantafyllou NI, Kapaki E, Papageorgiou S, Sgouropoulos P, Vassilopoulos D. A normative study of the Trail Making Test A and B in Greek adults. *Clin Neuropsychol* 2008;22:842–850.
35. Boone KB, Victor TL, Wen J, Razani J, Pontón M. The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Arch Clin Neuropsychol* 2007;22:355–365.
36. Manly JJ, Miller SW, Heaton RK, Byrd D, Reilly J, Velasquez RJ, Saccuzzo DP, Grant I. The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. The HIV Neurobehavioral Research Center (HNRC) Group. *J Int Neuropsychol Soc* 1998;4:291–302.
37. Manly JJ, Smith C, Crystal HR, Richardson J, Golub ET, Greenblatt R, Robison E, Martin EM, Young M. Relationship of ethnicity, age, education, and reading level to speed and executive function among HIV+ and HIV– women: The Women’s Interagency HIV Study (WIHS) Neurocognitive Substudy. *J Clin Exp Neuropsychol* 2011;33:853–863.
38. Weissberger GH, Salmon DP, Bondi MW, Gollan TH. Which neuropsychological tests predict progression to Alzheimer’s disease in Hispanics? *Neuropsychology* 2013;27:343–355.
39. Kemmotsu N, Enobi Y, Murphy C. Performance of older Japanese American adults on selected cognitive instruments. *J Int Neuropsychol Soc* 2013;19:773–781.
40. Buré-Reyes A, Hidalgo-Ruzzante N, Vilar-López R, Gontier J, Sánchez L, Pérez-García M, Puente AE. Neuropsychological test performance of Spanish speakers: Is performance different across different Spanish-speaking subgroups? *J Clin Exp Neuropsychol* 2013;35:404–412.
41. Fernández AL, Marcopulos BA. A comparison of normative data for the Trail Making Test from several countries: Equivalence of norms and considerations for interpretation. *Scand J Psychol* 2008;49:239–246.

42. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271–276.
43. Dilling H, Mombour W, Schmidt MH (Hrsg.). Internationale Klassifikation psychischer Störungen: ICD-10 (7. Aufl.), Kapitel V (F). Klinisch-diagnostische Leitlinien. Bern: Huber, 2010.
44. Lobo A, Launer LJ, Fratiglioni L, Andersen K, Di Carlo A, Breteler MM, Copeland JR, Dartigues JF, Jagger C, Martinez-Lage J, Soininen H, Hofman A. Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* 2000;54:S4–9.
45. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. *Lancet* 2006;368:387–403.
46. Bloudek LM, Spackman DE, Blankenburg M, Sullivan SD. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. *J Alzheimers Dis* 2011;26:627–645.
47. Peña-Casanova J, Sánchez-Benavides G, de Sola S, Manero-Borrás RM, Casals-Coll M. Neuropsychology of Alzheimer's Disease. *Arch Med Res* 2012;43:686–693.
48. Jahn T. Neuropsychologie der Demenz. In: Neuropsychologie psychischer Störungen (2. Auflage). Berlin, Heidelberg: Springer, 2010.
49. Perry RJ, Hodges JR. Attention and executive deficits in Alzheimer's disease: A critical review. *Brain* 1999;122:383–404.
50. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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* 1984;34:939–944.
51. Stokholm J, Vogel A, Gade A, Waldemar G. Heterogeneity in executive impairment in patients with very mild Alzheimer's disease. *Dement Geriatr Cogn Disord* 2006;22:54–59.
52. Steenland K, Macneil J, Bartell S, Lah J. Analyses of diagnostic patterns at 30 Alzheimer's disease centers in the US. *Neuroepidemiology* 2010;35:19–27.
53. Maruta C, Guerreiro M, de Mendonca A, Hort J, Scheltens P. The use of neuropsychological tests across Europe: The need for a consensus in the use of assessment tools for dementia. *Eur J Neurol* 2011;18:279–285.
54. Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST, Ganguli M. Patterns of cognitive decline in presymptomatic Alzheimer disease: A prospective community study. *Arch Gen Psychiatry* 2001;58:853–858.
55. Amieva H, Lafont S, Auriacombe S, Rainville C, Orgogozo JM, Dartigues JF, Fabrigoule C. Analysis of error types in the trail making test evidences an inhibitory deficit in dementia of the Alzheimer type. *J Clin Exp Neuropsychol* 1998;20:280–285.

56. Brown P, Devanand D, Liu X, Caccappolo E. Functional impairment in elderly patients with mild cognitive impairment and mild Alzheimer disease. *Arch Gen Psychiatry* 2011;68:617–626.
57. Van der Vlies AE, Verwey NA, Bouwman FH, Blankenstein MA, Klein M, Scheltens P, van der Flier WM. CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. *Neurology* 2009;72:1056–1061.
58. Shindo A, Terada S, Sato S, Ikeda C, Nagao S, Oshima E, Yokota O, Uchitomi Y. Trail making test part A and brain perfusion imaging in mild Alzheimer's disease. *Dement Geriatr Cogn Disord Extra* 2013;3:202–211.
59. Terada S, Sato S, Nagao S, Ikeda C, Shindo A, Hayashi S, Oshima E, Yokota O, Uchitomi Y. Trail making test B and brain perfusion imaging in mild cognitive impairment and mild Alzheimer's disease. *Psychiatry Res* 2013;213:249–255.
60. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, de Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B. Mild cognitive impairment. *Lancet* 2006;367:1262–1270.
61. Nordlund A, Rolstad S, Hellström P, Sjögren M, Hansen S, Wallin A. The Goteborg MCI study: Mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatry* 2005;76:1485–1490.
62. Arnáiz E, Almkvist O. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. *Acta Neurol Scand* 2003;107:34–41.
63. Saxton J, Snitz BE, Lopez OL, Ives DG, Dunn LO, Fitzpatrick A, Carlson MC, DeKosky ST. Functional and cognitive criteria produce different rates of mild cognitive impairment and conversion to dementia. *J Neurol Neurosurg Psychiatry* 2009;80:737–743.
64. Griffith HR, Netson KL, Harrell LE, Zamrini EY, Brockington JC, Marson DC. Amnesic mild cognitive impairment: Diagnostic outcomes and clinical prediction over a two-year time period. *J Int Neuropsychol Soc* 2006;12:166–175.
65. Chen N-C, Chang C-C, Lin K-N, Huang C-W, Chang W-N, Chang Y-T, Chen C, Yeh Y-C, Wang P-N. Patterns of executive dysfunction in amnesic mild cognitive impairment. *Int Psychogeriatr* 2013;25:1181–1189.
66. Alichniewicz KK, Brunner F, Klünemann HH, Greenlee MW. Structural and functional neural correlates of visuospatial information processing in normal aging and amnesic mild cognitive impairment. *Neurobiol Aging* 2012;33:2782–2797.
67. Pa J, Boxer A, Chao LL, Gazzaley A, Freeman K, Kramer J, Miller BL, Weiner MW, Neuhaus J, Johnson JK. Clinical-neuroimaging characteristics of dysexecutive mild cognitive impairment. *Ann Neurol* 2009;65:414–423.
68. Zanetti M, Ballabio C, Abbate C, Cutaia C, Vergani C, Bergamaschini L. Mild cognitive impairment subtypes and vascular dementia in community-dwelling elderly people: A 3-year follow-up study. *J Am Geriatr Soc* 2006;54:580–586.

69. Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR Jr, Feldman HH, Bokde ALW, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H, North American Alzheimer's Disease Neuroimaging Initiative (ADNI). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* 2012;33:1203–1214.
70. Perri R, Serra L, Carlesimo GA, Caltagirone C, Early Diagnosis Group of Italian Interdisciplinary Network on Alzheimer's Disease. Preclinical dementia: an Italian multicentre study on amnesic mild cognitive impairment. *Dement Geriatr Cogn Disord* 2007;23:289–300.
71. Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE, Alzheimer's Disease Neuroimaging Initiative. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Arch Gen Psychiatry* 2011;68:961–969.
72. D'Elia LF, Satz P, Uchiyama C, White T. *Color Trails Test. Professional Manual*. Odessa, Florida: Psychological Assessment Resources, 1996.
73. Lu L, Bigler ED. Performance on original and a Chinese version of Trail Making Test Part B: A normative bilingual sample. *Appl Neuropsychol* 2000;7:243–246.
74. Razzak RA. A Preliminary Study on the Trail-Making Test in Arabic–English Bilingual Young Adults. *Appl Neuropsychol Adult* 2013;20:53–60.
75. Axelrod BN, Aharon-Peretz J, Tomer R, Fisher T. Creating interpretation guidelines for the Hebrew trail making test. *Appl Neuropsychol* 2000;7:186–188.
76. Hashimoto R, Meguro K, Lee E, Kasai M, Ishii H, Yamaguchi S. Effect of age and education on the Trail Making Test and determination of normative data for Japanese elderly people: The Tajiri Project. *Psychiatry Clin Neurosci* 2006;60:422–428.
77. Seo EH, Lee DY, Kim KW, Lee JH, Jhoo JH, Youn JC, Choo IH, Ha J, Woo JI. A normative study of the Trail Making Test in Korean elders. *Int J Geriatr Psychiatry* 2006;21:844–852.
78. Cangoz B, Karakoc E, Selekler K. Trail Making Test: Normative data for Turkish elderly population by age, sex and education. *J Neurol Sci* 2009;283:73–78.
79. Lu L, Bigler ED. Normative data on Trail Making Test for neurologically normal, Chinese-speaking adults. *Appl Neuropsychol* 2002;9:219–225.
80. Wang T-Y, Kuo Y-C, Ma H-I, Lee C-C, Pai M-C. Validation of the Route Map Recall Test for Getting Lost Behavior in Alzheimer's Disease Patients. *Arch Clin Neuropsychol* 2012;27:781–789.
81. Maj M, D'Elia L, Satz P, Janssen R, Zaudig M, Uchiyama C, Starace F, Galderisi S, Chervinsky A. Evaluation of two new neuropsychological tests designed to minimize cultural bias in the assessment of HIV-1 seropositive persons: A WHO study. *Arch Clin Neuropsychol* 1993;8:123–135.

82. Agranovich AV, Panter AT, Puente AE, Touradji P. The culture of time in neuropsychological assessment: Exploring the effects of culture-specific time attitudes on timed test performance in Russian and American samples. *J Int Neuropsychol Soc* 2011;17:692–701.
83. Fasfous AF, Puente AE, Pérez-Marfil MN, Cruz-Quintana F, Peralta-Ramirez I, Pérez-García M. Is the Color Trails culture free? *Arch Clin Neuropsychol* 2013;28:743–749.
84. Kenny RA, Coen RF, Frewen J, Donoghue OA, Cronin H, Savva GM. Normative values of cognitive and physical function in older adults: Findings from the Irish Longitudinal Study on Ageing. *J Am Geriatr Soc* 2013;61 Suppl 2:S279–290.
85. McGuinness B, Barrett SL, Craig D, Lawson J, Passmore AP. Attention deficits in Alzheimer's disease and vascular dementia. *J Neurol Neurosurg Psychiatry* 2010;81:157–159.
86. Lin Y-C, Shih Y-C, Tseng W-YI, Chu Y-H, Wu M-T, Chen T-F, Tang P-F, Chiu M-J. Cingulum correlates of cognitive functions in patients with mild cognitive impairment and early alzheimer's disease: A diffusion spectrum imaging study. *Brain Topogr* 2014; 5: 393–402.
87. Dugbartey AT, Townes BD, Mahurin RK. Equivalence of the Color Trails Test and Trail Making Test in nonnative English-speakers. *Arch Clin Neuropsychol* 2000;15:425–431.
88. Mok N, Tsang L, Lee TMC, Llorente AM. The impact of language on the equivalence of Trail Making Tests: Findings from three pediatric cohorts with different language dominance. *Appl Neuropsychol* 2008;15:123–130.
89. Lee TM, Chan CC. Are Trail Making and Color Trails tests of equivalent constructs? *J Clin Exp Neuropsychol* 2000;22:529–534.
90. Lee TM, Cheung CC, Chan JK, Chan CC. Trail Making across languages. *J Clin Exp Neuropsychol* 2000;22:772–778.
91. Zhao Q, Guo Q, Li F, Zhou Y, Wang B, Hong Z. The Shape Trail Test: Application of a new variant of the Trail Making Test. *PLoS ONE* 2013;8:e57333.
92. Lu J-C, Guo Q-H, Hong Z. Trail Making Test used by Chinese elderly patients with mild cognitive impairment and mild alzheimer' dementia. *Chin J Clin Psychol* 2006;14:118–121.
93. Wang Q, Li W, Mao WL, Liu J, Zhao H, Liu C. 连线测验(中文修订版)在早期识别无痴呆型血管性认知障碍中的作用. *Chin J Gerontol* 2012;32:2018–2020.
94. Kim HJ, Baek MJ, Kim S. Alternative type of the Trail Making Test in nonnative English-speakers: The Trail Making Test-Black & White. *PLoS ONE* 2014;9:e89078.
95. Barncord SW, Wanlass RL. The Symbol Trail Making Test: Test development and utility as a measure of cognitive impairment. *Appl Neuropsychol* 2001;8:99–103.

96. Amodio P, Wenin H, Piccolo FD, Mapelli D, Montagnese S, Pellegrini A, Musto C, Gatta A, Umiltà C. Variability of Trail Making Test, Symbol Digit Test and Line Trait Test in normal people. A normative study taking into account age-dependent decline and sociobiological variables. *Aging Clin Exp Res* 2002;14:117–131.
97. Vanhanen M, Koivisto K, Kuusisto J, Mykkänen L, Helkala E-L, Hänninen T, Riekkinen P, Soininen H, Laakso M. Cognitive function in an elderly population with persistent impaired glucose tolerance. *Diabetes Care* 1998;21:398–402.
98. Kramer JH, Nelson A, Johnson JK, Yaffe K, Glenn S, Rosen HJ, Miller BL. Multiple cognitive deficits in amnesic mild cognitive impairment. *Dement Geriatr Cogn Disord* 2006;22:306–311.
99. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198.
100. Heinzl S, Schulte S, Onken J, Duong Q-L, Riemer TG, Heinz A, Kathmann N, Rapp MA. Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2014;146–173.
101. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials* 1981;2:93–113.
102. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C. The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer’s disease. *Neurology* 1989;39:1159–1165.
103. Bortz PDJ, Schuster PDC. Korrelation. In: *Statistik für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer, 2010: 153–182.
104. Lee IA, Preacher KJ. Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. (Accessed 02/08/2014 at <http://quantpsy.org>).
105. Ivarsson A, Andersen MB, Johnson U, Lindwall M. To adjust or not adjust: Nonparametric effect sizes, confidence intervals, and real-world meaning. *Psychol Sport Exerc* 2013;14:97–102.
106. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–1293.
107. Jonker C, Geerlings MI, Schmand B. Are memory complaints predictive for dementia? A review of clinical and population-based studies. *Int J Geriatr Psychiatry* 2000;15:983–991.
108. Kosunen O, Soininen H, Paljärvi L, Heinonen O, Talasniemi S, Riekkinen PJ. Diagnostic accuracy of alzheimer’s disease: A neuropathological study. *Acta Neuropathol* 1996;91:185–193.
109. Moosbrugger H, Kelava A. Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In: *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). Berlin Heidelberg, Springer, 2012: 7–26.

110. Bornstein RA, Baker GB, Douglass AB. Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *J Nerv Ment Dis* 1987;175:229–232.
111. Dikmen SS, Heaton RK, Grant I, Temkin NR. Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J Int Neuropsychol Soc* 1999;5:346–356.
112. Matarazzo JD, Wiens AN, Matarazzo RG, Goldstein SG. Psychometric and clinical test-retest reliability of the Halstead impairment index in a sample of healthy, young, normal men. *J Nerv Ment Dis* 1974;158:37–49.
113. Bauer L, Pozehl B, Hertzog M, Johnson J, Zimmerman L, Filipi M. A brief neuropsychological battery for use in the chronic heart failure population. *Eur J Cardiovasc Nurs* 2012;11:223–230.
114. Woods SP, Childers M, Ellis RJ, Guarnan S, Grant I, Heaton RK. A battery approach for measuring neuropsychological change. *Arch Clin Neuropsychol* 2006;21:83–89.
115. Giovagnoli AR, Pesce MD, Mascheroni S, Simoncelli M, Laiacona M, Capitani E. Trail Making Test: Normative values from 287 normal adult controls. *Ital J Neurol Sci* 1996;17:305–309.
116. Charter RA. A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *J Gen Psychol* 2003;130:290–304.
117. Miner T, Ferraro FR. The role of speed of processing, inhibitory mechanisms, and presentation order in trail-making test performance. *Brain Cogn* 1998;38:246–253.
118. Taylor R. Order effects within the Trail Making and Stroop Tests in patients with neurologic disorders. *J Clin Exp Neuropsychol* 1998;20:750–754.
119. Rosselli M, Ardila A. The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain Cogn* 2003;52:326–333.
120. Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and White elders. *J Int Neuropsychol Soc* 2002;8:341–348.
121. Bridges AJ, Holler KA. How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychol* 2007;13:528–538.
122. Wagner S, Helmreich I, Dahmen N, Lieb K, Tadic A. Reliability of three alternate forms of the Trail Making Tests A and B. *Arch Clin Neuropsychol* 2011;26:314–321.

## **Eidesstattliche Versicherung**

Ich, Thomas Günther Riemer, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „*Der Berliner Shape Trail Test – Charakterisierung und Validierung eines neuen Trail Making Tests*“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -[www.icmje.org](http://www.icmje.org)) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s. o.) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.

Datum

Unterschrift

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

## Publikationen

Dinges M, Riemer T, Schubert T, Prüss H. Musical hallucinations after pontine ischemia: The auditory Charles Bonnet syndrome? *J Neurol* 2013;260:2678–2680.

Heinzel S, Riemer TG, Schulte S, Onken J, Heinz A, Rapp MA. Catechol-O-methyltransferase (COMT) genotype affects age-related changes in plasticity in working memory: A pilot study. *BioMed Res Int* 2014;2014:414351.

Heinzel S, Schulte S, Onken J, Duong Q-L, Riemer TG, Heinz A, Kathmann N, Rapp MA. Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2014;146–173.

Schoofs N, Riemer T, Bald LK, Heinz A, Gallinat J, BERPohl F, Gutwinski S. Methadone and levomethadone - dosage and side effects. *Psychiatr Prax* 2014;41:82–87.

## **Danksagung**

Mein Dank gilt zunächst meinem Doktorvater Herrn PD Dr. Olaf Schulte-Herbrüggen für die konsequente, kritische und dabei stets motivierende fachliche Unterstützung in allen Abschnitten meiner Dissertation.

Meinen Kolleginnen Frau Andrea Lohse, Frau Janna Beckmann und meinem Kollegen Herrn Dr. Stephan Heinzl danke ich für die Unterstützung bei der Rekrutierung und Datenerhebung der Probanden. Im gleichen Sinne dankbar bin ich auch allen Mitarbeitern des Gerontopsychiatrischen Zentrums des St. Hedwig-Krankenhauses für die gute Zusammenarbeit bei der Akquise der Patienten.

Mein besonderer Dank gilt Herrn Dr. Benedikt Reuter, Herrn Dr. Alexander Riemer und Herrn Björn Elsner für das kritische Lesen meiner Dissertation und die daraus resultierenden wertvollen Hinweise für die inhaltliche und sprachliche Gestaltung der Arbeit.

Selbstverständlich danke ich allen Probanden und Patienten, die an der Studie teilgenommen haben und ohne deren Teilnahme diese Arbeit nicht möglich gewesen wäre.